NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

PREDICTING ENLISTED REENLISTMENT RATES
by
Arjay Nelson
September 2010
Thesis Advisor: Rachel Johnson Second Reader: Kevin Maher

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.			
1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** September 2010	**3. REPORT TYPE AND DATES COVERED** Master's Thesis	
4. TITLE AND SUBTITLE Predicting Enlisted Reenlistment Rates			**5. FUNDING NUMBERS**
6. AUTHOR(S) Arjay J. Nelson			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) BUPERS-34 5720 Integrity Drive Millington, TN 38054			**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB Protocol number ____N/A______.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited			**12b. DISTRIBUTION CODE**

13. ABSTRACT (maximum 200 words)

Manpower management and retention has been an issue for the military since the military became an all-volunteer force in 1973. Annually, the Bureau of Personnel Metrics and Analytics Branch (BUPERS-34) predicts Navy reenlistment rates and sets numeric reenlistment goals for the upcoming fiscal year. These goals ultimately take into account end strength considerations as well as Enlisted Community Manager requirements. BUPERS-34 uses linear regression to forecast what the expected reenlistment rate will be, given current conditions; if no force shaping actions (e.g., reduce accessions, change personnel policies) are taken. If the forecasted reenlistment rate is different than requirements from an end strength/community management perspective, then the force shapers in the Manpower, Personnel, Training and Education Policy Division (N13) have a signal that steps may need to be taken to bring the two in line. In this thesis, the current BUPERS-34 Navy reenlistment prediction method is evaluated and alternative models to improve the prediction accuracy are suggested. Results of the analysis suggest the removal of several variables from the current model, due to lack of statistical significance, and the addition of Selected Reenlistment Bonus as a predictive variable for reenlistment.

14. SUBJECT TERMS Forecasting Reenlistment Rates, Manpower Modeling, Enlisted Modeling, Prediction Models, Predicting Reenlistment Rates, Regression Analysis, Design of Experiment.			**15. NUMBER OF PAGES** 87
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540-01-280-5500 Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

PREDICTING ENLISTED REENLISTMENT RATES

Arjay J. Nelson
Commander, United States Navy
B.A., University of California at Santa Barbara, 1991

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
September 2010

Author: Arjay J. Nelson

Approved by: Rachel T. Johnson
Thesis Advisor

Kevin J. Maher
Second Reader

Robert F. Dell
Chairman, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Manpower management and retention has been an issue for the military since the military became an all-volunteer force in 1973. Annually, the Bureau of Personnel Metrics and Analytics Branch (BUPERS-34) predicts Navy reenlistment rates and sets numeric reenlistment goals for the upcoming fiscal year. These goals ultimately take into account end strength considerations as well as Enlisted Community Manager requirements. BUPERS-34 uses linear regression to forecast what the expected reenlistment rate will be, given current conditions; if no force shaping actions (e.g., reduce accessions, change personnel policies) are taken. If the forecasted reenlistment rate is different than requirements from an end strength/community management perspective, then the force shapers in the Manpower, Personnel, Training and Education Policy Division (N13) have a signal that steps may need to be taken to bring the two in line. In this thesis, the current BUPERS-34 Navy reenlistment prediction method is evaluated and alternative models to improve the prediction accuracy are suggested. Results of the analysis suggest the removal of several variables from the current model, due to lack of statistical significance, and the addition of Selected Reenlistment Bonus as a predictive variable for reenlistment.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

2FI	Two-Factor Interaction
ACOL	Annualized Cost of Leaving
ARI	Army Research Institute
BLS	Bureau of Labor Statistics
BUPERS	Bureau of Naval Personnel
BUPERS-34	Bureau of Naval Personnel Metrics and Analytics Branch
CBO	Congressional Budget Office
CMF	Career Management Field
CNA	Center for Naval Analysis
CNP	Chief of Naval Personnel
CY	Calendar Year
DCNO	Deputy Chief of Naval Operations
DMDC	Defense Manpower Data Center
DTIC	Defense Technical Information Center
DoD	Department of Defense
DOX	Design of Experiment
EAOS	Expiration of Active Obligated Service
ECM	Enlisted Community Manager
EMF	Enlisted Master File
EMR	Enlisted Master Record
EPA	Enlisted Program Authorizations
FY	Fiscal Year
GAO	Government Accounting Office
HYT	High Year Tenure
IMA	Integrated Moving Average
LOS	Length of Service
LTE	Long Term Extensions
MOS	Military Occupation Specialty
N13	Manpower, Personnel, Training & Education Policy Division

N1	Deputy Chief of Naval Operations of Manpower, Personnel, Training & Education
NAVADMIN	Naval Administrative Message
NAVRET	Navy Enlisted Retention Statistics Reporting System
NEC	Naval Enlisted Classification
NID $(0, \sigma^2)$	Normally and Independently Distributed With a Mean of Zero and a Constant Variance
NRMS	Navy Retention Monitoring System
NPC	Naval Personnel Command
NSIPS	Navy Standard Integrated Personnel System
OBLISERVE	Obligated Service
OPNAV	Office of Chief of Naval Operations
OPTEMPO	Operational Tempo
PCS	Permanent Change of Station
PRD	Projected Rotation Date
PVCOL	Present Value Cost of Leaving
RE	Reenlistments
RMC	Regular Military Compensation
SCOL	Stochastic Cost of Leaving
SDAP	Special Duty Assignment Pay
SECNAV	Secretary of the Navy
SRB	Selected Reenlistment Bonus
UE	Unemployment Rate
WebRMS	Web-based Retention Monitoring System

EXECUTIVE SUMMARY

Manpower management and retention has been an issue for the military since the inception of the all-volunteer force in 1973. A large body of research has been conducted to define, measure, and discover contributing factors related to retention and attrition. The Bureau of Naval Personnel Metrics and Analytics Branch (BUPERS-34) uses multivariate linear regression to fit models that predict the upcoming fiscal year reenlistment for specific enlisted zones. This thesis focuses on three specific enlistment zones: A, B, and C, which are based on completed years of service. While the current regression models were originally based on sound research, the models have become somewhat outdated and are in need of evaluation. In this thesis, the current BUPERS-34 Navy reenlistment prediction method is evaluated and alternative models to improve the prediction accuracy are suggested.

Three main problems are identified with the current reenlistment rate regression models. First, the current models potentially violate the mathematical assumptions that the models are based on. Second, the models are shown to contain insignificant variables. Finally, several of the variables in the current models require predictions as inputs in order to make forecasts for future values, thus creating additional noise in the forecasts.

This thesis uses several statistical techniques to evaluate the current problems with the forecasting models and recommends alternative models. The models suggested are more robust than the current BUPERS-34 prediction models and provide improved forecasts with lower prediction variability. The alternative models eliminate insignificant variables, improve model fit, and incorporate additional compensation (e.g., Selective Retention Bonus) that effect zone reenlistment rate predictions.

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

There are a number of people I would like to acknowledge without whom I would have had difficulty completing this thesis. A well-deserved thank you goes to Dr. Rachel Johnson for her unending patience as my advisor. She tirelessly answered questions, provided guidance, and gave feedback through numerous edits.

I would also like to thank those military and civilian members on the BUPERS-34 and N13 staffs that were willing to meet with me and/or give me information that was valuable throughout my thesis process. They get my deepest thanks. They provided the data, the BUPERS-34 Reenlistment Prediction model and its significance, and excellent directions on how to replicate and use the model, as well as, answering a multitude of questions along the way.

Lastly, I thank my family, who spent more time without me than any of us would have liked. To Renee Nelson, my wonderful and supportive wife, you have my heartfelt appreciation for understanding why I was away at school throughout the week, and when home, studying on the weekends and sacrificing our family time. Thank you for your support these last 28 months. I could not have done it without you, so this degree is yours to share. To Hunter and Kendall, this degree is meaningless without your love and support. Thank you for your understanding when I was not available, and for your excitement when I came home on the weekends.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

Annually, the Bureau of Personnel Metrics and Analytics Branch (BUPERS-34) predicts Navy reenlistment rates and numeric reenlistment goals as part of establishing the following fiscal year retention goals. The BUPERS-34 forecasts what the expected reenlistment rate will be, given current conditions, if no force shaping actions are taken (e.g., reduce accessions, change personnel policies) to change the expected behavior of sailors. If the forecasted reenlistment rate is different from the rate required to meet end strength, then the force shapers in the Manpower, Personnel, Training, and Education Policy Division (N13) have a signal that steps may need to be taken to bring the two in line.

This thesis analyzes the current BUPERS-34 Navy Reenlistment Rate Prediction model and considers alternative methods that improve the accuracy and validity of the model.

A. PURPOSE

The Chief of Naval Personnel (CNP) is a three-star admiral in charge of Navy's manpower readiness. Dual-titled, the CNP also serves as Deputy Chief of Naval Operations (Manpower, Personnel, Training & Education) and oversees the Bureau of Naval Personnel (BUPERS), Navy Personnel Command, and the Navy Manpower Analysis Center. As one of four Deputy Chiefs of Naval Operations (DCNO) (Figure 1), with the identification of N1, the DCNO performs all strategy and resource policies and serves as a single resource sponsor for all manpower and training program matters (Navy.mil, 2007). The N1 also performs all Capitol Hill related duties, including all Congressional testimony for matters pertaining to the Manpower, Personnel, Training, & Education command.

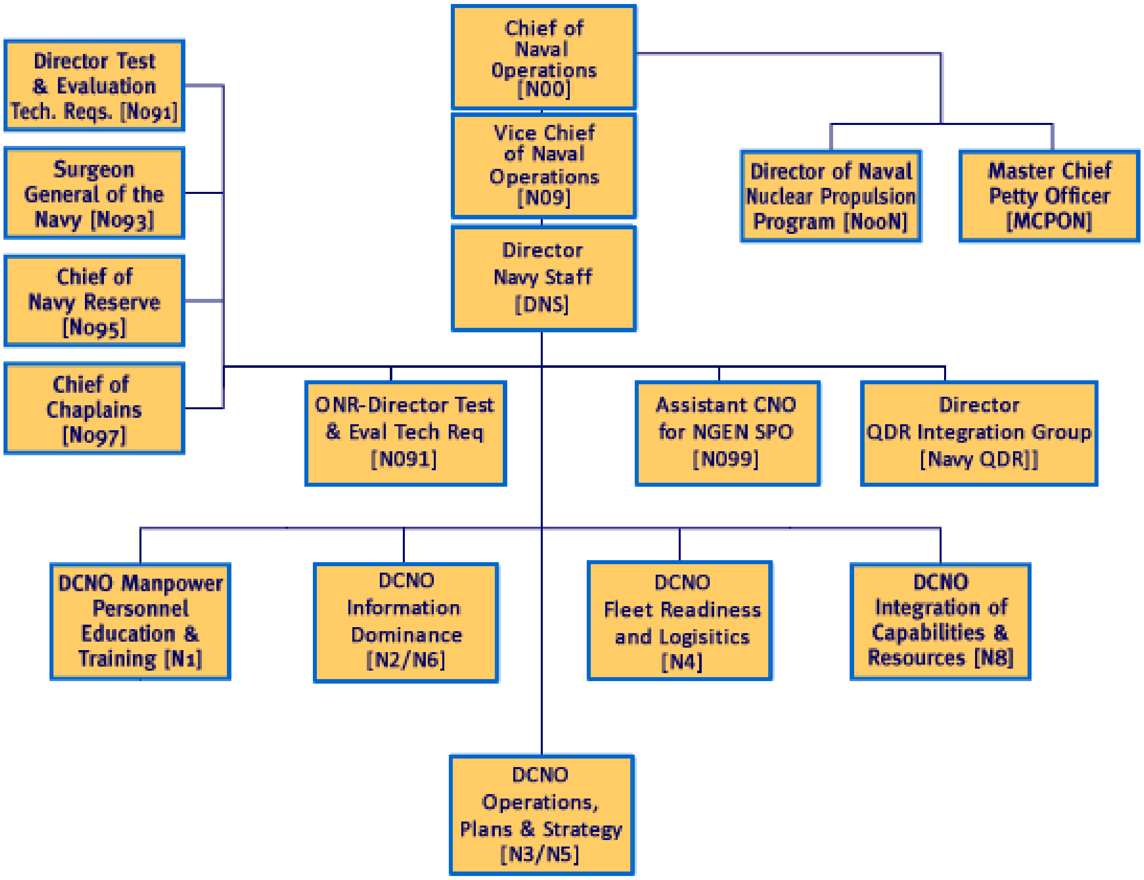


Figure 1. Chief of Naval Operations Organizational Chart (From Navy.mil, 2010)

Each fiscal year (FY), the N1 establishes reenlistment goals to best position the navy to meet end strength requirements, while responding to likely factors that will shape Navy's retention efforts. End strength requirements are fiscal year military personnel authorizations given by Congress under Title 10, United States Code (Defense Technical Information Center [DTIC], 2009). The National Defense Authorization Act prescribes the number of personnel authorized. This number usually changes each FY based on budget and personnel requirements. The requirement is that the end strength obligation is met on 30 September each FY. For FY 2010, the Secretary of Defense requested from Congress specific service personnel authorizations as recommended by the respective service secretaries. Navy end strength received authorization for 328,800 active duty personnel (See Table 1). Subsequently, in order for Navy to meet the congressional end strength authorization for each fiscal year, Navy determines their reenlistment goals by reenlistment zone (as defined and explained in the next paragraph) and releases those goals in a Navy Administrative Message (NAVADMIN).

	FY 2009		FY 2010	Change from	
Service	Authorized	Request	Committee Recommendation	FY2009 Authorized	FY2010 Request
Army	532,400	547,400	547,400	0	15,000
Navy	326,323	328,800	328,800	0	2,477
USMC	194,000	202,100	202,100	0	8,100
Air Force	317,050	331,700	331,700	0	14,650
DoD	1,369,773	1,410,000	1,410,000	0	40,227

Table 1. FY 2010 Military Personnel Authorizations (From DTIC.mil, 2010)

1. FY 2009 Retention Message

NAVADMIN 348/08 (Ferguson, 2008b) and 333/09 (Ferguson, 2009) updated the definition of reenlistment zones and standardized enlisted retention measures of effectiveness for all zones. Enlistment zones are specific length of service (LOS) parameters (Table 2) used to set Navy retention goals. The zones are shown in Table 2.

Zone A	Less than 6 years of service (YOS)
Zone B	6 to less than 10 YOS
Zone C	10 to less than 14 YOS
Zone D	14 to 20 YOS
Zone E	Greater than 20 YOS

Table 2. Reenlistment Zones

Based on the zones shown in Table 2, NAVADMIN 348/08 summarized the Navy's attainment of FY-08 retention objectives (Table 3). Both reenlistment goal and actual values are presented in Table 3. The Navy exceeded retention goals established for FY08 across zone A and B and was short in zone C. Actual numeric reenlistment rates exceeded their goal by 310 reenlistments resulting in 26,510 total reenlistments compared to a goal of 26,200. Strong command and leadership attributed to the navy attaining 101 percent of their total numeric reenlistments for sailors in zones A through C (Ferguson, 2008b).

Zone	Goal	Actual
ZONE A (0 TO 6 YEARS OF SERVICE)	48 PERCENT	50.7 PERCENT
ZONE B (6 TO 10 YEARS OF SERVICE)	58 PERCENT	59.8 PERCENT
ZONE C (10 TO 14 YEARS OF SERVICE)	82 PERCENT	80.2 PERCENT

Table 3. All Navy FY-08 Reenlistment Rate (From Ferguson, 2008b)

The BUPERS-34 reenlistment rate prediction and numeric reenlistment goals for zones A, B, and C have a great impact on the ability for the navy to sustain targeted manpower and readiness. Good predictions can assist in reducing personnel overages or underages, and subsequent costs associated with missing the target end strength. Any improvement in BUPERS-34 ability to predict reenlistment, as discussed later, may result in a greater manpower cost savings and/or readiness state.

The following narrative best describes leadership's desired direction for achieving Navy retention goals:

> Because we are becoming smaller, with more demands and a wider range of missions, the Navy must continue to shape the force to achieve the best "fit." "Fit" means having a trained sailor, at the right place, at the right time. Achieving fit through retention means moving beyond the aggregate reenlistment rate goals towards meeting retention requirements based on rating and length of service. Individual goals are essential in influencing the desired reenlistment behavior for our most critical ratings. (Ferguson, 2008b, p. 1)

Reenlistments are important to delivering target end strength. A large part in planning for the future of the Navy is related to predicting how many sailors there will be each year. Enlistments and reenlistments are part of this planning. BUPERS-34 utilizes simple forecasting tools in order to predict future reenlistments. This thesis evaluates the predicative capability of the Zone A through C reenlistment models and suggests methods for improvement. Improved predictions can ultimately result in cost savings for the Navy.

B. MOTIVATION FOR THESIS

The quote “all models are wrong, some are useful" by George Box (Box & Draper, 1987), the 20th century statistician, is a well-known quote in statistics and may best describe the challenge behind evaluating the BUPERS-34 reenlistment rate prediction model and necessity to review and update the model.

BUPERS-34 reenlistment rate predictions are aggregate rates. Their FY09 reenlistment rate predictions for zones A, B, and C on first glance (refer to Table 2 for zone descriptions and Table 3 for FY09 predictions), appear to be relatively close to the actual rates. On average, the FY09 predictions when compared to actual reenlistments overshoot by approximately two percentage points for all three zones, which is significant with a large number of reenlistments. However, measuring BUPERS-34 real prediction accuracy is much more challenging because the prediction serves as a baseline to implement “levers” at the beginning of the FY. These levers, or manpower retention actions (e.g., selected reenlistment bonus, approving or disapproving waivers), continually drive reenlistment rates as close to the respective FY numeric manpower goals per zone by reevaluating the levers in meeting targeted monthly goals. In May 2009, Rear Admiral (RADM) Holloway, Manpower, Personnel, Training & Education Policy Division (N13) said, “we review each rating weekly with the community managers and take a monthly look at how we are looking with re-enlistments before making adjustments. We’re carefully watching all re-enlistment and retention behavior- we don’t want to get caught flat footed” (Faram, 2010, p.30).

Overshooting FY-09 goals by one percent, or approximately 310 sailors, is costly. Using a conservative example, in 2006, Table 4 shows that the Congressional Budget Office (CBO) estimated that the regular military compensation for a single E-5 with six years experience was approximately $45,000. Subsequently, the total cost to the Navy for overshooting their FY-09 manpower goals by 310 sailors most likely exceeded $14 million (310 x $45,000 [2006 dollars]). This figure does not account for any bonuses, special pays, or other non-cash or deferred benefits such as retirement pay and health care that would increase total compensation to approximately $100,900 per sailor, a cost of over $31 million to the Navy. Additionally, overshooting retention goals does not take into account unnecessary bonuses (overpaying to stay).

(2006 Dollars)									
Pay Grade	E-1	E-2	E-3	E-4	E-5	E-6	E-7	E-8	E-9
Typical Age	18	19	20	22	25	31	37	40	44
Average Years of Experience	<2	<2	<2	3	6	12	18	21	25
Compensation: Enlisted (Single)									
Cash	29,700	32,000	32,900	37,200	**45,000**	54,000	63,400	72,400	85,900
Noncash and deferred cash	25,300	26,900	27,600	31,200	35,600	41,800	48,500	54,300	64,900
Total	**54,900**	**58,900**	**60,500**	**68,400**	**80,600**	**95,700**	**111,900**	**126,600**	**150,700**
Compensation: Enlisted (Married with children)									
Cash	32,800	34,700	36,300	40,400	47,200	56,800	65,200	72,800	89,600
Noncash and deferred cash	37,300	38,900	39,700	49,200	53,700	59,800	64,800	70,200	81,100
Total	**70,100**	**73,600**	**76,000**	**89,700**	**100,900**	**116,600**	**130,000**	**143,000**	**170,700**

Table 4. Estimated Compensation for Enlisted Personnel (From CBO, 2007)

Undershooting is also severe because of the potential impact to the loss of readiness and ability to meet mission. Under estimating goals has costs that are more difficult to measure because the remedy may result in over compensation (e.g., overcompensating sailors to stay or return), low morale (e.g., increased operations tempo due to manpower shortages), or poor personnel fit (e.g., retaining the wrong sailors) to meet mission.

The financial and/or readiness cost to the Navy for overshooting or undershooting their reenlistment rate and retention goals is significant. Improving the accuracy and validity of the current prediction model will minimize these costs and inefficiencies to attain the target goals. However, it is challenging to measure the accuracy of the BUPERS-34 Reenlistment Rate Prediction model. This is because the model predicts zone reenlistment rate behavior prior to the next FY before many retention levers or force shaping actions (e.g., bonus levels, Perform to Serve monthly retention boards, and high year tenure waiver approvals or disapprovals) are implemented or withdrawn as needed to attain the targeted end strength by the end of the that FY. This makes the original reenlistment rate predictions difficult to evaluate on their own because they are “fitted,” and, therefore, the BUPERS-34 prediction accuracies are open for interpretation.

An evaluation and validation of the reenlistment rate model and numeric retention goals is appropriate and justified in an ever changing and dynamic environment. This thesis assesses the reliability and robustness of the BUPERS-34 Reenlistment Rate Prediction model to meet targeted retention goals, and proposes a new and improved model.

C. PROBLEM STATEMENT AND THESIS OUTLINE

The current multi-variate linear regression model developed and used by BUPERS-34 to predict reenlistment rates for zones A, B, and C, is analyzed in this thesis. Recommendations for changes in the model that improve accuracy and precision of the predictions are made. Chapter II provides a literature review that investigates previous studies regarding retention models and discusses different approaches regarding enlisted behavior. Chapter III discusses the Navy Retention Monitoring System (NRMS) database that is used for retention analysis and describes the BUPERS-34 Reenlistment Rate Prediction model. Chapter IV evaluates the current reenlistment rate prediction model used by BUPERS-34. Chapter V discusses new proposed prediction models. Chapter VI analyzes the subsequent data output, derives a conclusion, and proposes follow-on research.

II. RELATED LITERATURE

Manpower management and retention has been an issue for the military since the military became an all volunteer force in 1973. A large amount of research has been conducted to define, measure, and discover contributing factors related to retention and attrition in qualitative and quantitative reports, studies, and papers. Much of the research contained in the literature makes great effort to explain the numerous factors contributing to retention.

In this thesis, the Navy's reenlistment prediction model is analyzed. Reenlistment and retention are sometimes used interchangeably, but do have a difference that should be discussed. Retention rates are the number of personnel retained out of a specified group of people. For example, retention rate can apply to the organization as a whole. Reenlistment rates, a subset of retention rates, refer to a specific group of people that are eligible for reenlistment during specified periods. The groups of people used for calculating reenlistment rates are, in general, those that have served their obligated length of service and have the option to either reenlist or leave the service. This thesis assumes that factors contributing to retention and reenlistment are somewhat similar, thus the literature review discusses models focusing on both retention and reenlistment.

This literature review focuses on two areas of related military manpower research and its effect on retention: military non-compensation retention models and compensation retention models. Non-compensation models are models that investigate the effects of non-compensation factors (e.g., variables) such as unemployment rate and operation tempo that may be significant to retention. Compensation models investigate the significance of military pay, civilian pay, bonuses and other forms of compensation that may be significant to retention. The current BUPERS-34 Reenlistment Rate Prediction model is a non-compensation model. This thesis investigates adjusting the model to include bonuses (e.g., Selective Reenlistment Bonus [SRB]) at the aggregate level and varying the periods of which the data is modeled to analyze and provide the statistical variation necessary to produce significant estimates to predict the reenlistment rate.

A. MILITARY NON-COMPENSATION RETENTION MODELS

The United States military has experienced a reduction in force since the end of the first Gulf War in 1991. Subsequently, non-compensation retention models and/or variables have been examined to see their effect on reenlistment retention. In the economics literature in particular, there is a focus on looking at a metric called elasticity. Elasticity is the ratio of the percent change in one variable to the percent change in another variable. For example, pay elasticity for reenlistment, measures the percent change in reenlistment associated with a 1-percent increase in pay.

Goldberg (1986) provides estimates of the effect of unemployment on enlisted retention. The Goldberg study looks at data from FY 1977 to FY 1984 where large swings in the unemployment rate make estimates of unemployment effects on retention more critical and provide the statistical variation necessary to produce significant estimates. A time series analysis was used to compare the effects of military pay and unemployment rate on retention rate. It resulted in the appearance that either variable had a significant effect on retention trends but the separate effects were impossible to distinguish. When rate specific SRBs were included in military pay, military pay was distinguishable from unemployment rate effects on retention. Unemployment was found to have a significant effect upon the reenlistment rate for seven of the nine rating groups studied, and a significant effect upon both the extension rate and the total retention rate for all nine rating groups. However, because the pay elasticities (which include SRBs) are three to five times as large as the unemployment elasticities (e.g., the percent change in reenlistment associated with a 1 percent increase in unemployment), the unemployment rate may be offset by much smaller percentage increases in military pay. This study reports the statistically significant effect of unemployment on retention. However, unemployment rate is of only secondary importance when compared to military pay.

Budding et al. (1992) concluded that retention models are sensitive to the specification of individual promotion opportunities at the end of their first term of enlistment. Expected time to E5 promotion has a significant effect on first-term retention in both the pay ratio and the annualized cost of leaving (ACOL) formulations of the retention model. Other things equal, a 10 percent promotion slowdown is associated with 14 percent and 8 percent reductions in Army and Air Force retention rates, respectively. The results show that traditional retention approaches have not been adequately controlled for promotion tempo and that promotion could be used to complement military pay and bonus policies in retaining quality personnel in hard-to-find-skills.

Hansen and Wenger (2003) examined the costs and benefits of retention as a way to develop rating-specific reenlistment goals for zone A enlisted personnel. Each rating identifies and quantifies the primary costs and benefits to the Navy of higher reenlistment. For example, if the benefits of higher reenlistment (e.g., retention of more experienced sailors, increased manpower) are greater than the costs, the cost-effective level of reenlistment is higher than its current level. The results indicate that economic conditions do affect the cost-effective level of reenlistment and that a deterioration of the civilian economy will generate higher retention without any need to increase reenlistment bonuses. Additionally, the study found that although the Navy still has to pay higher seniority costs from increased retention, the value of the additional experience, combined with recruiting and training cost savings, outweighs the cost of the higher reenlistment rate. In contrast, improvements in economic conditions act like a "tax" on SRB effectiveness. For some ratings, it is cost-effective to raise SRBs to offset the impact of economic conditions. For other ratings, however, it would be prohibitively expensive to return reenlistment rates to previous levels.

Questor, Hattaingadi and Shuford (2006) examined the effect of U.S. Marine Corps deployment tempo on Marine reenlistment behavior in FY-04. They find that first-term Marines making reenlistment decisions in FY-04 who deployed to a crises area and spend more total days deployed than their peers have lower reenlistment rates. Additionally, they find that deployment tempo negatively affects Marines without dependents most significantly. The study results indicates no relationship between days deployed and reenlistment decisions for second and third term Marines, and officers.

B. MILITARY COMPENSATION RETENTION MODELS

Concern about the retention of active-duty military personnel prompted numerous proposals to improve military pay and benefits in the 1980s. Several enlisted retention models were implemented and/or considered by the armed services.

To measure the effect of changes in military compensation on reenlistment decisions, the Congressional Budget Office (CBO) developed a military retention model. The CBO military retention model predicts the effects on retention of future compensation changes by assuming that reenlistment decisions are motivated by military and civilian compensation over an individual's entire remaining career (CBO, 1981). The model is formulated using a weighted average of future pays, called "perceived pay," where the weights are both discount rates and the person's probability of remaining in the military. This model captures the effects only of the largest compensation components (i.e., regular military compensation, SRBs, and retirement pay). It asserts that retention decisions are motivated by compensation over an individual's entire remaining career, and that a pay change over the entire future pay stream should exert a strong effect on junior personnel. This study is a technical description of the CBO retention model that has been used for several senate and congressional reports prior to 1981. The study does not offer any recommendations; it concludes that CBO retention model over-predicts enlisted retention rates because of incorrect military pay assumptions for two reasons. First, by including only monetary values, it ignores such intangible, but critical, factors as an individual's preference, or "taste," for military service (CBO, 1981). Second, it ignores the effect of past compensation practices (e.g., higher SRBs) that may lead to whether an individual reenlists again. Warner (1981) conducted an analysis on four major models for predicting the effect of military pay on retention; the Present Value of the Cost of Leaving (PVCOL) model, the Annualized Cost of Leaving (ACOL) model, the Stochastic Cost of Leaving (SCOL) model, and the Air Force- Congressional Budget Office model. All of these models are similar in that they attempt to measure military pay relative to civilian pay, and “taste” (e.g., likes and dislikes) for staying in the military. They differ in their income stream (cost of leaving) to remain in the military for one more term and the income stream to leave. The cost of leaving is then related to the retention rate. The ACOL and SCOL are more descriptively accurate than earlier models, because they measure “taste” for military service and provide more sensible predictions than earlier models. The PVCOL model does not measure military to civilian compensation differences and the Air Force- CBO model over-predicts reenlistment rates.

The Center for Naval Analyses developed two models for projecting enlisted end strength in 1981: the Prophet model, and the ACOL model. The Prophet model tracks the distribution of the force by years of remaining obligated service, but does not allow reenlistment rates to vary in response to changes in compensation. Reenlistment rates are estimated by length of service. Conversely, the ACOL model does allow reenlistment rates to vary in response to changes in compensation where the reenlistment rate is estimated by the effects of compensation on reenlistment but does not track the distribution of the force by years of remaining obligated service. Goldberg and Hagar (1981) compared the career force projections of these models to actual historical experience over the period FY 78–FY 80. They found that the ACOL projections are more accurate than the Prophet projections and, subsequently, adjusting reenlistment rates in response to pay changes is more important than tracking the force by years of remaining obligated service.

Trumble and Flanagan conducted a study in 1990 for the Navy Personnel Research and Development Center (NPRDC) that reviewed existing forecasting and simulation methodologies to improve forecasts of naval officer retention rates. Two major types of models were compared, ACOL, which was the official forecasting model used by the Office of the Assistant Secretary of Defense, the Navy, and the Air Force to provide personnel loss rate forecasts at various levels of disaggregation, and Dynamic Retention (DR) models. Both models were discussed in detail with respect to the ability to model and evaluate manpower policies of interest to NPRDC staff. The DR model was considered the best theoretically because it was able to adequately capture the dynamic effects of a temporary pay changes. The DR model does so with detailed modeling of an officer's entire service career with an underlying "taste for the service" parameter. However, the formulation and implementation of the DR model was more costly than the ACOL model and required significant improvements, resulting in the ACOL model continued usage (Trumble, 1990).

Goldberg (2001) provides a survey on enlisted retention models from 1973 to 2001 and offers some analysis and recommendation for future work. Goldberg’s survey review is extensive and summarizes the influence of many retention models and modeling techniques. The survey’s primary focus of enlisted retention models begins with the impact of the ACOL model and its influence on other models and statistical techniques from the mid-1970s to 1990s. The survey then reviews pay elasticity models, the retention effects of other variables that are not pay related (i.e., length of deployment, incidence of sea duty, and percentage of time spent away while not deployed) and the effect of a SRB on those variables. The paper attempts to decompose the variation of pay elasticities (e.g., sensitivity analysis in regards to compensation) in terms of data handling (e.g., treatment of enlisted ineligibles and extensions), modeling technique, and elasticity composition. Goldberg asserts that many pay elasticity models used to forecast retention use different techniques resulting in great variations in their forecasts. He concludes by recommending a “controlled experiment” to eliminate any confounding differences between the variations of several pay elasticity models in order to develop a more precise model.

The United States Army Research Institute for the Behavioral Sciences (ARI) (2005) conducted an analysis on the significance of SRBs on enlisted retention by including SRBs into the ACOL model to estimate the financial incentive to stay. The model was generated using logistic regression. ARI measured the effects of SRBs on reenlistments, at zone A (between 17 months and 6 years of active service), Zone B (between 6 and 10 years of active service), and Zone C (between 10 and 14 years of active service) at three levels of occupational aggregation. The three level are all-Army (i.e., Army as a whole), career management field (CMF), and military occupation specialty (MOS). The results for Zone A at all levels of occupational aggregation indicate that reenlistment bonuses have a positive and statistically significant effect on Zone A reenlistments. The magnitude of the effect varied by occupation, but a one-level increase in SRB at Zone A typically increases the reenlistment rate by three to seven percentage points, depending upon the occupation. The results for Zone B are significant at both the CMF and MOS levels. Results for Zone C, where reenlistment rates are typically very high, are similar but not as good as the Zone A and B results. Additionally, Zone C sometimes relies on higher-level occupational aggregations to obtain estimates.

C. SUMMARY

As reviewed, military compensation models, such as the ACOL model, have been modified many times since their inception in the 1970s to analyze their effect on retention. Through their many modifications (e.g., SRB inclusion), they continue to remain useful. ACOL models, in particular, have remained influential models used by military analysts as a measurement of pay elasticity and a sailors “taste” to stay in the military (Goldberg, 2001). As well, non-compensation models and/or variables (e.g., promotion tempo, unemployment rate, and economic conditions) have proven useful to measure retention behavior; however, several studies imply (e.g., Goldberg, 1986) that econometric and/or compensation variables (e.g., ACOL, SRB) have greater significance in measuring the variability in military enlisted retention models.

The purpose of this review is to provide insight into the many methods, models, and strategies used to predict reenlistment rates. Predicting a sailor’s reenlistment rate is very complex because there is not one dominant method or strategy to model retention. Additionally, a sailor’s behavior is nearly impossible to predict due to a dynamic and ever changing environment. Retention variables and models need continuous analysis and modifications.

This thesis uses the insights from the related literature as a reference to explore improvements to the BUPERS-34 reenlistment rate model’s methodology and variable selection.

III. BUPERS-34 METRICS AND ANALYTICS BRANCH

BUPERS-34 has the responsibility to monitor, analyze, predict and report enlisted retention and attrition trends. Through N1, their prediction and trend analysis provides annual (and monthly) enlisted retention targets (goals) and trends to the Fleet and other Echelon II commanders.

Retention measures (e.g., reenlistment rates) are calculated in the NRMS, Navy's authoritative source of retention (Ferguson, 2008a). This chapter introduces the NRMS, discusses the calculation of the Navy's reenlistment model, and provides an example of the use of the reenlistment model for a particular fiscal year.

A. NAVY RETENTION MONITORING SYSTEM

The Navy Retention Monitoring System (NRMS) is a web-based application developed in 2004. It combines the legacy Web based Retention Monitoring System (WebRMS) and Navy Enlisted Retention Statistics Reporting System (NAVRET) to provide timely and accurate reporting and analysis of reenlistment, retention, and attrition data. NRMS expands on the functionality of NAVRET and WebRMS to enhance the capability to provide effective and efficient reporting and analytical information for staff, program managers, decision makers, and fleet units. In addition to information available in WebRMS, data contained in the Navy Standard Integrated Personnel System (NSIPS) are incorporated into an Enterprise Data Warehouse, from where all NRMS report information is drawn (SPAWAR, 2004).

1. Access and Deliverability

NAVRET, which was based on a Microsoft Access database, has several drawbacks. These drawbacks are: (1) NAVRET is accessed by all users using just a single password; (2) it is not available to most Command Career Counselors; (3) all historical data has to be downloaded to the local user's computer; (4) and it does not meet updated security requirements (SPAWAR, 2004). Subsequently, NRMS has improved security requirements meeting all federal and the Freedom of Information Act and information security requirements. The Bureau of Naval Personnel Metric and Analysis Branch (BUPERS-34) administers the system and user accessibility.

Additionally, NRMS partitions and restricts data and personal information to three user levels:

a. The Chief of Naval Operations (CNO) N13, Manpower, Personnel, Training & Education Policy Division, can access all NRMS reports and has full Ad hoc capability within the NRMS Data Mart (Enterprise Data Warehouse). Ad hoc capability is available for all subordinate commands based on the Administrative Unit Identification Code Tree. N13 is able to view the full Social Security Number (SSN) of all members.

b. Career Counselor Level 1 is composed of Center for Career Development (CCD) members, all Fleet and Force Counselors, and other individuals as defined by the CCD. For comparison purposes, these users have the ability to view all delivered reports. Ad hoc capability is available for the user’s command and his/her subordinate commands. All reports in this level display only the last six digits of a member’s SSN.

c. Career Counselor Level 2 includes those users assigned as Command Career Counselors at the unit or command level. Career Counselor Level 2 users are able to view only NAVRET based NRMS delivered reports. All reports at this level display only the last six digits of a member’s Social Security Number (SSN). Reports are limited to the last three years of data.

2. Functionality

NRMS provides access to over 10,000 registered navy personnel that may retrieve personnel data from 1992 to present for retention reporting using business intelligence capabilities (Welgan, 2010). Business intelligence capabilities are functions that build quantitative processes for a business, or in this case, the Navy to arrive at optimal decisions and to perform analytical computations within NRMS and its populated database. These capabilities in the business world frequently involve data mining, statistical analysis, predictive analytics, predictive modeling, and business process modeling. However, NRMS has not fully incorporated all of these analytical and predictive capabilities. Instead, NRMS is used most often for the "measurement" component of the NRMS business intelligence capability. The measurement program creates a hierarchy of performance metrics and benchmarking that informs users (Navy leadership, Community Managers, and Command Career Counselors [CCCs]) about progress towards retention goals.

Navy manpower specialists (N13), BUPERS-34, Community Managers, and Fleet and Force Counselors monitor reenlistment, retention, and attrition trends in numerous categories and monitor the effectiveness of Command Retention Programs of subordinate commands.

CCCs use NRMS to monitor their command's reenlistment, retention, and attrition data in a variety of modes to provide the Commanding Officer and the Command Retention Team the information needed to establish and maintain an effective Career Information/Retention programs.

3. Report Types

Ad hoc reporting is available. These are reports that allow Navy manpower specialists (N13) and Career Counselor Level 1 users to create reports to gather information that are not covered by NRMS Corporate Reports to support analysis. A module, called Business Objects Universe Report, allows the user to generate Ad hoc queries. Users will interact with data using representations of information, or “Business Objects,” with which they are familiar. Data elements are grouped into folders of logical collections referred to as “classes.” Ad hoc reporting is based on classes of personnel data elements residing in the NRMS Data Mart (SPAWAR, 2004).

In general, the most widely used reports are “Corporate Reports.” Corporate Reports are prepared reports by the administrator (BUPERS-34) that require no additional user manipulation. The standard Corporate reports are the 12 Month Cumulative; FYTD (Fiscal Year to Date); and Monthly Reenlistment, Retention, and Attrition Reports.

4. NRMS Calculations and Modeling Support

Retention measures, predefined calculations and standards within NRMS, are used within Corporate and Ad Hoc Reporting. BUPERS-34 uses some of these measures to predict reenlistment rates though regression analysis. However, the BUPERS-34 Reenlistment Rate Prediction model is not calculated within NRMS. NRMS serves to support the model by providing the critical data.

The following two sections serve as examples of naval personnel reenlistment variables (e.g., dimensions) and retention measures that are available within NRMS.

a. NRMS Dimensions

Dimensions variables allow NRMS to sort data in numerous ways to modify, narrow, or expand the scope of NRMS reports. Table 5 presents a sample of dimensions in NRMS.

Dimension Panel	Description
Armed Forces Qualification Test (AFQT)	The panel allows you to query AFQT scores by category (e.g., CAT 1) and then by score.
Members (Branch)	The members panel allows you to select USN (active duty), USNR (reservist), or both for your report.
Length of Service (LOS)	The LOS panel allows you to sort by Zone.
Number of Months	This panel allows you to sort by a specific time period (e.g., FY to date, 12-month cumulative)

Table 5. NRMS Sample Dimension Panels

b. Standard Retention Measures in NRMS

"Measures" are various calculations that NRMS can perform. Table 6 lists a sample of the most commonly used Navy standard retention measures, and their definitions and computations for active duty personnel as defined in NAVADMIN 333/08 (Ferguson, 2008a).

Measure	Definitions and Computations
Attrition	Enlisted personnel lost from the Navy prior to their expiration of active obligated service (EAOS).
Attrition Rate	The proportion of sailors who leave active duty prior to reaching their EAOS. Measures Non-EAOS loss behavior.
Attrition Rate Computation	(Non-EAOS Losses) / (Non-EAOS Inventory)
Long Term Extension (LTE)	Extension of service greater than 24 months
Non-EOAS Inventory	Includes all sailors in a particular zone who are greater than 90 days from their EAOS.
Reenlistment (RE)	Formal reenlistment greater than 24 months
RE Rate	Measures EAOS behavior
RE Rate Computation	(RE + LTE) / (RE + LTE + EAOS losses)

Table 6. Standard Retention Measures

The NRMS is a significant improvement over previous computer based retention monitoring systems. It offers users and administrators (e.g., BUPERS-34) secure and efficient means to obtain and evaluate retention data over the web. Additionally, NRMS is scalable and has the potential to expand its capabilities to provide more analytical functions and data for modeling retention behavior.

B. BUPERS-34 REENLISTMENT RATE MODEL

Using the data pulled from the NRMS database, BUPERS-34 predicts out-year (i.e., next FY) reenlistment rates and reenlistment numbers using a multi-variate linear regression prediction method (BUPERS-34, 2009). The general multiple linear regression equation is (Montgomery, 2006):

$$y= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon$$

Customarily x_i is called the independent (predictor or regressor) variable, y is called the dependent (response) variable, and ε is the statistical error. The β is the model coefficient (regression slope) and β_0 is the intercept, which are fit through the least squares method, and that minimize the sum of the squares of the errors. Є are the errors and are assumed to be normally and independently distributed with a mean of zero and a constant variance (NID $[0, \sigma^2]$).

BUPERS-34 has developed separate prediction models for reenlistment zones A, B, and C. These zones are considered the most significant to maintain operational readiness. The FY2010 BUPERS-34 Multiple Linear Regression response and predictor variables for zones A, B, and C are shown in Table 7. The current BUPERS-34 Reenlistment Rate Model predicts the reenlistment rate at the organization level (Navy aggregate) vice at the unit (e.g., command, squadron) or enlisted rating level (e.g., Aviation Technician, Personnel Man).

Variable	Variable Description
	Zone A
y	**Reenlistment Rate**. Reenlistment rate data from the previous 11 FYs is obtained from NRMS.
x_1	**End Strength**. Change in zone A end-strength from the previous 11 FYs is obtained from NRMS.
x_2	**Unemployment Rate**. Unemployment rate data from the previous 11 calendar years (CY) is obtained from the Bureau of Labor Statistics.
x_3	**Attrition Rate**. Attrition rate data from the previous 11 FYs is obtained from NRMS.
	Zone B
y	**Reenlistment Rate**. Reenlistment rate data from the previous 15 FYs is obtained from NRMS.
x_1	**End Strength**. Total end-strength at the start of the FY for previous 15 FYs is obtained from NRMS.
x_2	**Unemployment Rate**. Unemployment rate data from the previous 15 calendar years (CY) is obtained from the Bureau of Labor Statistics.
x_3	**Attrition Rate**. Attrition rate data from the previous 15 FYs is obtained from NRMS.
	Zone C
y	**Reenlistment Rate**. Reenlistment rate data from the previous 15 FYs is obtained from NRMS.
x_1	**End-Strength**. End-strength data at the start of the FY for sailors with 10-13 years LOS is obtained for the previous 15 FYs from NRMS.
x_2	**Unemployment Rate**. Unemployment rate data from the previous 15 calendar years (CY) is obtained from the Bureau of Labor Statistics.
x_3	**Attrition Rate**. Attrition rate data from the previous 15 FYs is obtained from NRMS.

Table 7. Zones A, B, and C Response and Predictor Variables

Each zone (A, B, and C), is individually modeled at the organization level in order to predict enlisted reenlistment rates for each zone. Reenlistment zones are consistent with SRB zones A, B, and C as defined in NAVADMIN 333/08 (Ferguson, 2008a). To remain consistent with the prescribed reenlistment zones, NRMS calculates and reports Navy retention measures, such as, reenlistment rate, attrition rate, and end strength, which are used in the BUPERS-34 model.

As observed in Table 7, the model variables for zone A, B, and C are extremely similar. Each of the models contains the variables “**Attrition Rate,**” “**Unemployment Rate,**” and “**End Strength.**” These three models differ in the way their respective end strength prediction variable is calculated. Zone A uses the last 11 FY years of respective variable data (i.e., reenlistment rate, attrition rate, end strength, and unemployment rate) and its end strength is computed by calculating the total numeric change in zone A end strength from the previous two FYs. For example, FY 2010 zone A change in end strength was 1,474 which was calculated by subtracting the total zone A end strengths for FY2009 (150,655) by FY2008 (149,181).

Zone B uses the last 15 years of respective variable data; **End Strength** is computed by the total Navy end strength at the start of the FY. Similar to zone B, zone C uses the last 15 years of data for its model. **End Strength** is calculated from the start of the FY for sailors with a length of service from 10 to 13 years.

Unemployment rates are derived from the Bureau of Labor Statistics. BUPERS-34 uses the total national unemployment rate from either the last 11 or 15 calendar years (vice fiscal years) for zone A, or B and C, respectively.

Depicted in Figure 2 is the BUPERS-34 FY 2010 multiple linear regression process for predicting reenlistment rates.

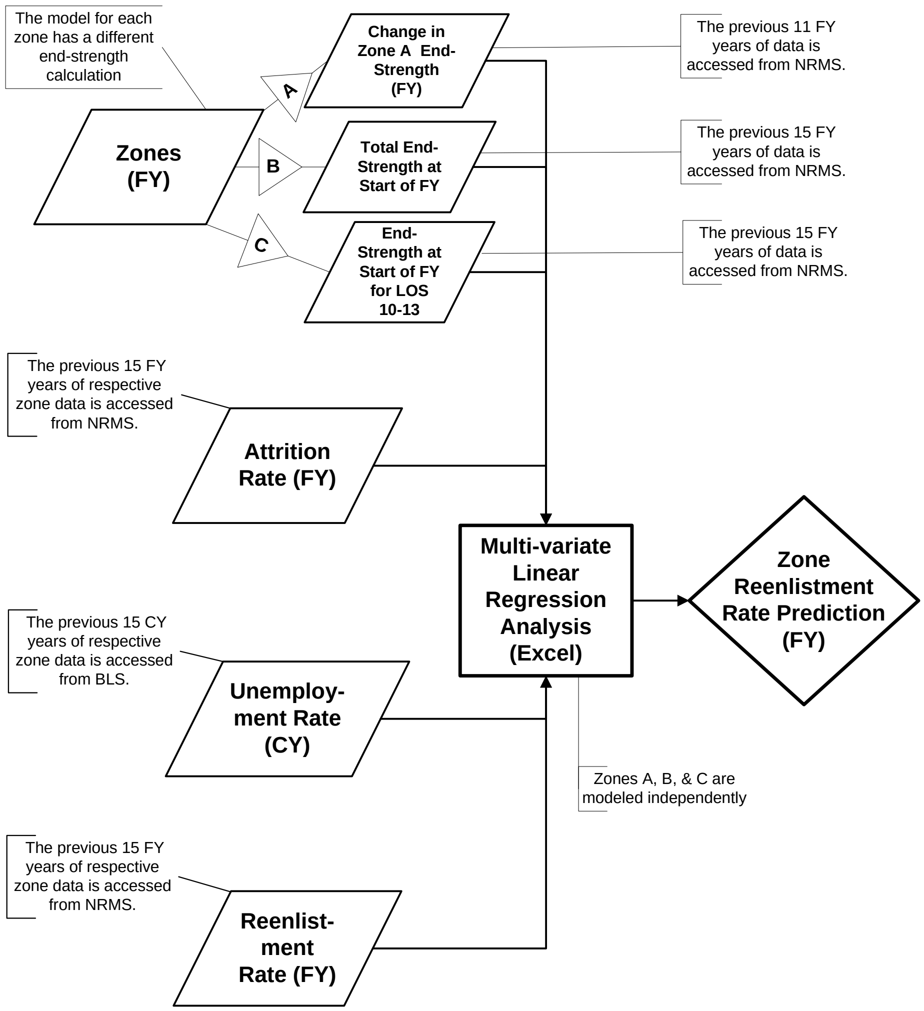


Figure 2. BUPERS-34 Zones A, B, and C Linear Regression Model Process For Predicting FY10 Reenlistment Rates

1. BUPERS-34 FY 2010 Zone A Regression Analysis Process and Prediction

Data is collected for the response and predicator variables for each zone from NRMS and the Bureau of Labor Statistic (BLS) to build a data set and perform regression analysis. To illustrate the BUPERS-34 regression analysis process and reenlistment rate prediction, zone A is used as an example.

The data for FY09-FY10 used to fit the zone A model is shown in Table 8.

Fiscal Year	Y Reenlistment Rate	X_1 Change in Zone A End-Strength	X_2 Unemployment Rate	X_3 Attrition Rate
1999	0.4755	2004	0.042	0.1341
2000	0.5141	7170	0.040	0.1289
2001	0.6005	10636	0.047	0.1089
2002	0.5885	9901	0.058	0.1015
2003	0.6021	8480	0.060	0.0829
2004	0.5081	-7173	0.055	0.0737
2005	0.5319	-2793	0.051	0.0779
2006	0.5149	-10048	0.046	0.0768
2007	0.4585	-10163	0.046	0.0840
2008	0.5061	-7292	0.058	0.0905
2009	0.5566	-1942	0.089	0.0843
2010	?	1474	0.094 (Estimate)	0.0721

Table 8. BUPERS-34 FY 2010 Zone A Data Set

Multiple linear regression is performed with the zone A data set using Excel resulting in the output shown in Table 9:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.958015072
R Square	0.917792878
Adjusted R Square	0.882561255
Standard Error	0.016737155
Observations	11

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.649646482	0.051207182	12.68663	4.37E-06
X_1 Change in Zone A End-Strength	6.68541E-06	8.30815E-07	8.046814	8.78E-05
X_2 Unemployment Rate	0.541799053	0.455838599	1.188577	0.273361
X_3 Attrition Rate	-1.53488826	0.357542556	-4.29288	0.003598

Table 9. BUPERS-34 FY 2010 Zone A Regression Analysis Results

Because there are only 11 observations (Figure 3), it is hard to see if there is a violation of NID. However, because this data is in a time series it is assumed that correlation among the data may exist. This is because the data collected is ordered by year and there may be trends in rates from one year to the next.

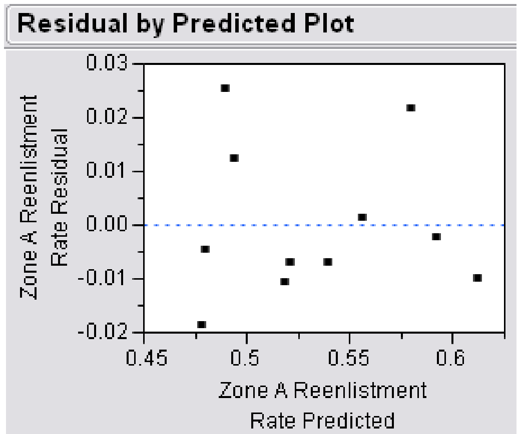


Figure 3. BUPERS-34 Zones A Residual by Predicted Plot

From the results in Table 9, the fitted regression equation can be written as:

$$Y_{zone\ A} = 0.649 + 0.000007x_1 + 0.541x_2 - 1.534x_3$$

Because BUPERS-34 is required to predict FY reenlistment rates in August of the preceding year, August and September values are estimated to derive a final FY value to be multiplied by their respective coefficient in the fitted regression equation (above). Subsequently, in order to use the linear regression equation as a forecasting tool to predict the zone reenlistment rates for FY10, the FY attrition rate and change in zone A end strength for FY2009 is partially estimated, and the unemployment rate for FY10 is predicted by a Department of the Navy economist (Chilson, personal communication, 2010).

For the FY10 reenlistment rate prediction, predicator variable data was obtained through NRMS up to August and estimated values were made from that data resulting in a FY year-end value resulting in a **Change in Zone A End-Strength** of 1474 sailors, and a **Attrition Rate** of 7.2 percent. The Department of the Navy predicted an **Unemployment Rate** for CY 2010 of 9.4 percent (Chilson, personal communication, 2010). A reenlistment rate of 59.5 percent was calculated from the following fitted equation:

$$.595 = 0.649 + 0.0000068*(1474) + 0.541*(.094) - 1.534*(.072)$$

2. BUPERS-34 FY 2010 Zone Reenlistment Rate

Table 10 summarizes BUPERS-34 FY 2010 reenlistment rate predictions for zones A, B, & C:

	ZONE A Reenlistment Rate	**ZONE B Reenlistment Rate**	**ZONE C Reenlistment Rate**
BUPERS 34 Prediction	**59.5 percent**	**69.5 percent**	**84.2 percent**

Table 10. BUPERS-34 FY 2010 Zone Reenlistment Rate (From Chilson, 2009)

C. FISCAL YEAR REENLISTMENT RATES AND NUMERIC RETENTION GOALS

Near the end of each FY, BUPERS-34, Enlisted Community Managers (ECM), End Strength planners (N104), and N13 convene as a working group to determine the next FY retention goals. The BUPERS-34 reenlistment rate predictions are used as reenlistment expectations for zones A, B, and C and are used to in identify the need for potential force shaping actions if goals and expectations diverge. The ECMs and end strength planners provide their recommendations for manpower requirements (e.g., enlisted rating needs) and end strength targets (e.g., total Navy personnel), respectively. N13 facilitates the working group’s process to resolve manpower requirements and end strength targets resulting in an enlisted retention goal recommendation that best balances enlisted rating needs with end-strength assumptions and the BUPERS-34 reenlistment rate prediction.

Near the end of FY 2009, the working group determines FY 2010 retention goals (Table 11).

	ZONE A Reenlistment Rate / Reenlistment Number	ZONE B Reenlistment Rate / Reenlistment Number	ZONE C Reenlistment Rate / Reenlistment Number
ECM Continuation Need	70 percent / 18,246	52 percent / 8,262	63 percent / 5,827
N104 End-Strength Assumptions	58 percent / 13,293	61.1 percent / 8,494	85.8 percent / 6,235
BUPERS 34 Prediction	59.5 percent/ 13,225	69.5 percent / 8,650	84.2 percent / 6,050
Recommendation	**59 percent / 13,500**	**60 percent / 8,300**	**71 percent / 5,800**

Table 11. FY10 Retention Goals (From Chilson, 2009)

The working group's recommendation is forwarded to N1 for approval. N1 modifies the recommendation as necessary to adjust to new data, insights and/or requirements since delivery the working group's recommendation.

In Figure 4, a flow of the retention process illustrates how the "All Navy FY 2010 Retention Goals" are determined and how the retention goals are resolved, approved, reported to the Secretary of the Navy and Congress, and distributed to the Fleet for implementation.

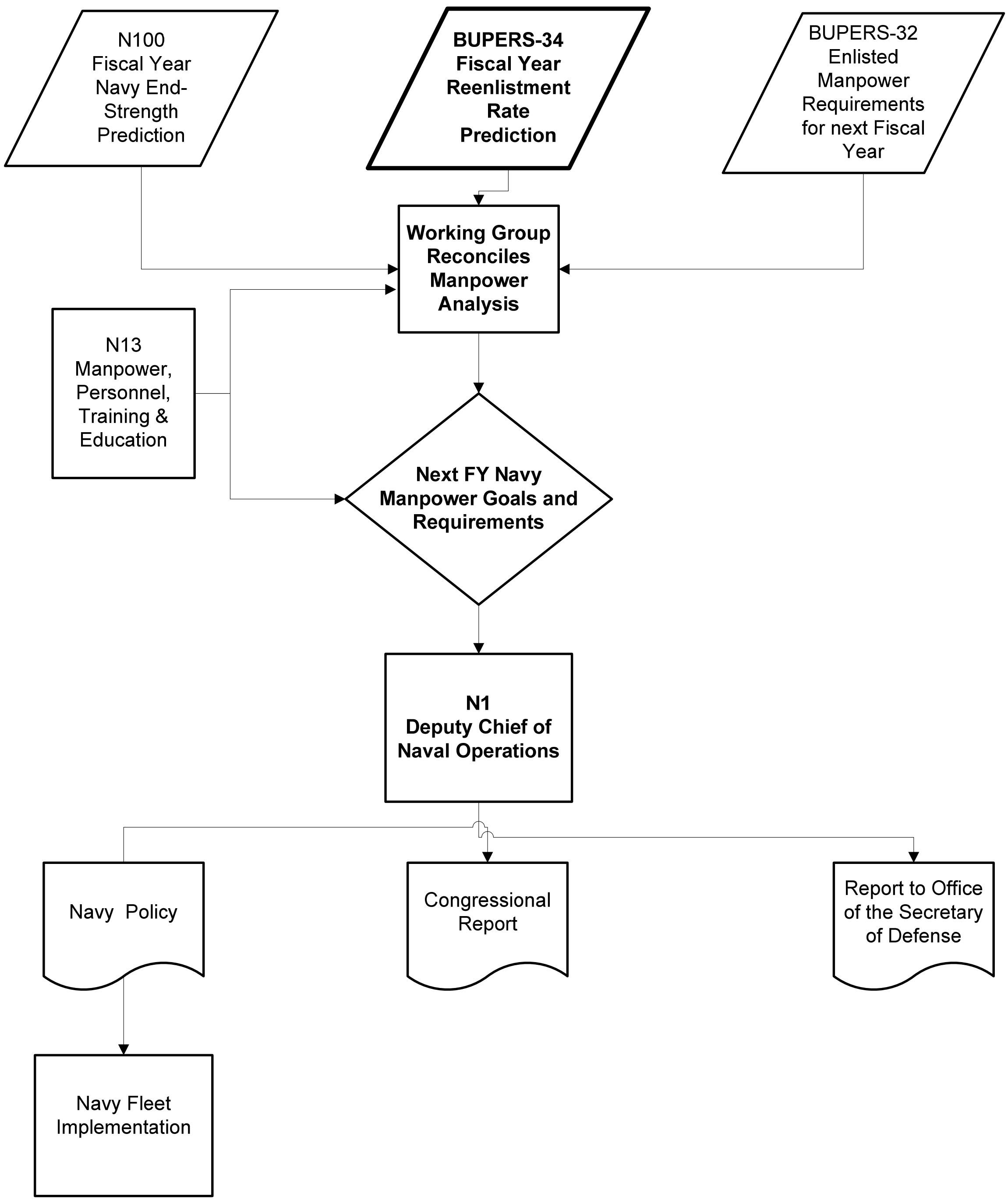


Figure 4. BUPERS-34 Reenlistment Rate Prediction and Reporting Process

D. SUMMARY

The BUPERS-34 Reenlistment Rate Prediction model predicts zone reenlistment rates for the succeeding FY at the aggregate level (i.e., Navy as a whole). The model uses non-compensation variables (Table 7) for zones A, B, and C. Their corresponding data is collected for the last 11-15 years from NRMS and the BLS based on data available for the respective zone, and then BUPERS-34 uses Excel to conduct multiple linear regression on the respective zone data to determine the coefficients for zones A, B, and C FY reenlistment rates (Figure 2). The resulting coefficients are multiplied by a predicted unemployment rate for the upcoming FY and the ending FY values for the current year's end strengths and attrition rates.

The resulting zone reenlistment rate predictions serve as a base line to assist in establishing Navy retention goals to meet end strength targets and manpower requirements for the upcoming FY (Table 10). BUPERS-34, BUPERS-32, N100, and N13 consolidate their information and reconcile their differences resulting in their retention goal recommendations (Table 11) being forwarded to N1 for final approval and made into policy (Figure 4).

The focus of this thesis is on evaluating the current reenlistment rate model and developing a plan for improving the predictive capability of the model. There are several problems with the current reenlistment rate model. The following descriptions illustrate the three main problems with the current model:

- Violation of Assumptions
  - The current model uses regression analysis to make predictions on time series data. The assumption used in linear regression is that the error (residuals) are NID(0, σ^2).
  - In some of the data, a strong correlation among the residuals can be observed. This violation in assumption will cause problems with the model results. For example, the variance may actually be higher than reported.

- Use of Insignificant Variables

    - The linear regression model was developed in FY08 using variables that today (and at that time) are no longer useful in the prediction.

    - Without continuously evaluating the fit of the model, there are variables that have become insignificant in terms of predicting reenlistment rate.

    - The use of insignificant variables in a model can cause over dispersion problems and lead to inadequate results

- Prediction within the Model

    - The linear regression model uses several variables to fit reenlistment rates for each zone A, B, and C. Some of these variables require predictions in order to make a forecast for future values of reenlistment rate, which makes the model difficult to use and adds more variability to the response.

The problems highlighted above provide additional motivation for the research in this thesis. Chapter IV presents an evaluation of the current model and Chapter V shows the development of several alternative models that could be used in place of the current prediction model.

IV. EVALUATION OF THE CURRENT REENLISTMENT RATE PREDICTION MODEL

Chapter III presented the NRMS database used to pull, sort, and store variables of interest to BUPERS-34. At the end of the chapter, several problems with the current model were highlighted. This chapter looks into the problems in further detail.

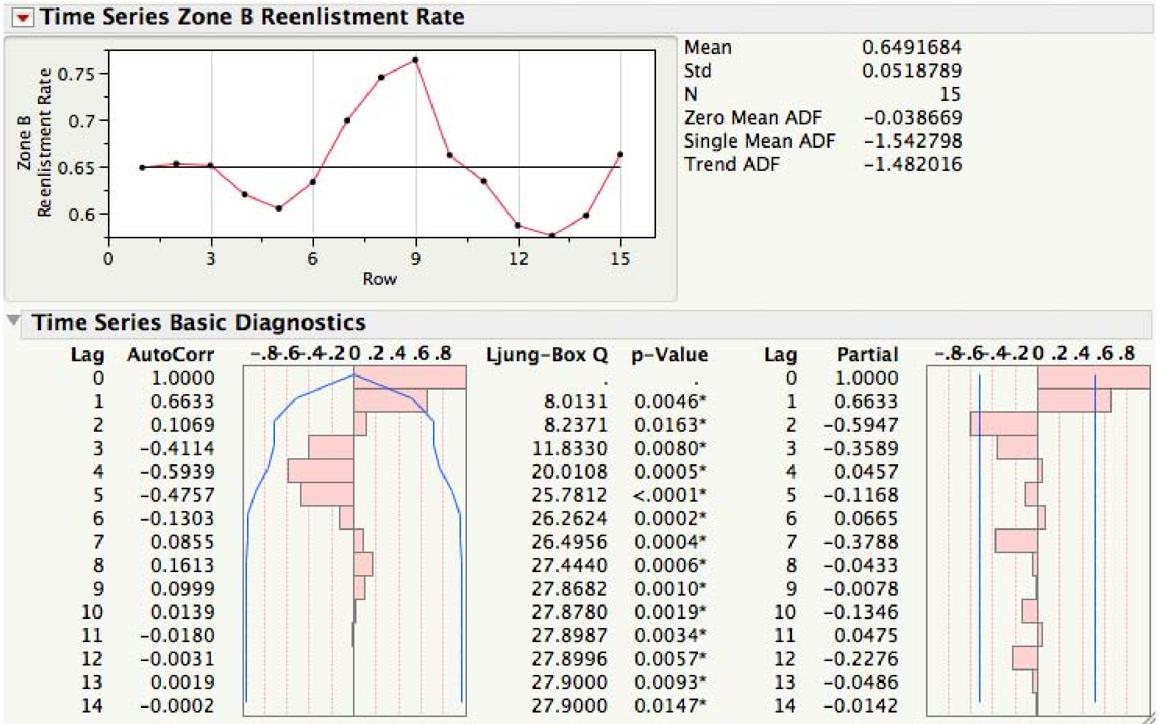
Section A in this chapter evaluates the current model. Specifically, the assumption of NID $(0, \sigma^2)$ residuals is evaluated and the significance of the current variables is studied. Section B presents a unique application of experimental design. The unique application of experimental design techniques is used to evaluate the use of data, both in terms of frequency of time slices and amount of historical data used, and also the use of variables to study the fit of the regression model to the data.

A. EVALUATING THE CURRENT BUPERS-34 MODEL FOR PREDICTING REENLISTMENT RATE

At the end of Chapter III, several problems with the current model are described. Two of those problems are the violation of assumptions for the linear model and the use of insignificant variables in the current model. Those problems are illustrated in this section.

1. Violation of Assumptions

Using linear regression for time series data is not always advisable because time series data can be significantly correlated. This will lead to a violation of the assumptions used for fitting least squares. An illustration of this violation is shown in Figure 5. Figure 5 presents the time series plot, autocorrelation function, and partial autocorrelation function for Zone B reenlistment rate data.



Mean	0.6491684
Std	0.0518789
N	15
Zero Mean ADF	-0.038669
Single Mean ADF	-1.542798
Trend ADF	-1.482016

Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	0.6633	8.0131	0.0046*	1	0.6633
2	0.1069	8.2371	0.0163*	2	-0.5947
3	-0.4114	11.8330	0.0080*	3	-0.3589
4	-0.5939	20.0108	0.0005*	4	0.0457
5	-0.4757	25.7812	<.0001*	5	-0.1168
6	-0.1303	26.2624	0.0002*	6	0.0665
7	0.0855	26.4956	0.0004*	7	-0.3788
8	0.1613	27.4440	0.0006*	8	-0.0433
9	0.0999	27.8682	0.0010*	9	-0.0078
10	0.0139	27.8780	0.0019*	10	-0.1346
11	-0.0180	27.8987	0.0034*	11	0.0475
12	-0.0031	27.8996	0.0057*	12	-0.2276
13	0.0019	27.9000	0.0093*	13	-0.0486
14	-0.0002	27.9000	0.0147*	14	-0.0142

Figure 5. Zone B Time Series Correlation

The plots in Figure 5 indicate that the Zone B reenlistment data is both autocorrelated and partially autocorrelated by the significant lags shown. The correlation between data one lag apart (one year in this data) is 0.663.

In regression analysis, the residuals are assumed to be normally and independently distributed. With such heavy dependencies in both the response data (reenlistment rate) and several of the inputs, there is an occasional violation of the independence assumption in the residuals. Time series analysis can be used to remove this correlation. In many of the regressions that are analyzed, the assumption of NID residuals is not violated. However, there are several instances of violation, such as the one pointed out in Figure 5. Based on the work in this thesis, the recommendation is to use time series analysis or perform a transformation on the response if a violation is detected.

2. Use of Insignificant Variables

The current BUPERS-34 model was developed and deployed for use in 2008. Changes in Navy manpower and personnel policies (e.g., end strength requirements, bonus levels) and the economy (e.g., increasing unemployment rate) have led to changes in reenlistment behavior. When using linear regression, it is important to evaluate the fit of the model. This includes determining whether independent variables in the model have a significant impact on the response (dependent) variable.

Table 12 shows the BUPERS-34 Reenlistment Rate Prediction adjusted R-squared, which givens an indication of model fit, and also shows the p-value for each of the variables in zones A, B, and C. An asterisk next to a p-value indicates that the variable is significant to the model at α= .05. Consequently, the unemployment rate is found to be insignificant to measuring the variability in zone A, B, & C reenlistment rate prediction models. However, BUPERS-34 includes unemployment rate in their prediction model for all three zones.

Zones	**Model Fit**	**Independent Variables (without interactions)**		
	Adjusted R-square	**End Strength**	**Unemployment Rate**	**Attrition Rate**
Zone A	**0.883**	0.0001*	0.2684	0.0035*
Zone B	**0.461**	0.0409*	0.0751	0.0145*
Zone C	**0.756**	0.0002*	0.0991	0.0000*

* P Value Significance at .05

Table 12. BUPERS-34 FY 2010 Reenlistment Rate Prediction Model Adjusted R-Squared and P-values for Each Variable

As indicated with the values in Table 12, unemployment rate is not statistically significant in fitting the linear regression model in the BUPERS-34 model. Consequently, the model is over-fitted because unemployment rate is not statistically significant at the targeted .05 significance level. This will result in poor predictive performance.

B. DESIGNING AN EXPERIMENT TO EVALUATE MODEL FIT

An experiment is a test or series of tests in which purposeful changes are made to the input variables of a process or system so that we may observe and identify the reasons for changes that may be observed in the output response (Montgomery, 2008). This thesis conducts an experimental design as the basis to determine the significant input variables in the BUPERS-34 Reenlistment Rate Prediction model by determining which variables are most influential on the response (output) variable (i.e., standard deviation and adjusted R-square of the fitted models). In this effort, a statistical design of experiments (e.g., factor screening, regression analysis) is used so that the appropriate data is collected and analyzed using statistical methods, resulting in objective conclusions.

Factor screening is used in this process to systematically vary input factors in order to identify those factors that produce a significant change in the response variables. Additionally, factor screening is used to estimate the magnitude and direction of individual factor effects as well as factor interaction effects on the response variable. In general, factor screening is best when conducted using only two levels of the factors. In this experiment a low-level and a high-level screening is used.

Multivariate linear regression is used to determine what factors in the screening experiment have a significant effect on the response. In a linear regression model the response variable (y), is related to predictor variables (x_i), through the following general equation:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{1,2} x_1 x_2 + \ldots + \beta_n x_n + \epsilon$$

The standard multivariate linear regression model tests the following hypothesis:

H_0 : $\beta_0 = \beta_1 = \beta_2 = \ldots = \beta_n = 0$; where n represents the number of coefficients

H_1 : At least one coefficient does not = 0

In order to gain insight into the construction and robustness of the BUPERS-34 Reenlistment Rate Prediction model and areas that can improve the model, an experiment is performed on the model to analyze different processes that can be conducted on the model that are affected minimally by external sources of variability. Standard deviation and adjusted R-square are the measurements used to determine and evaluate which process is best.

This design of experiments (DOX) seeks to analyze the strength and effects of the variables in the current BUPERS-34 Navy reenlistment forecasting method and improve the performance of the model and/or consider alternative models for improvement. At the end of the previous section, the presence of insignificant variables in the current model is discussed. In the following subsections, a design of experiments is used to systematically test the influence of inclusion of model terms, amount of data used, and period of data on the fitted regression models produced.

1. Selection of the Response Variables

In this experiment, there are two response variables (Y_1, Y_2), standard deviation and adjusted R-square. Standard deviation measures how closely the model fits the data. Thus, with lower standard deviation the model is assured to more accurately represent reenlistment rate. Subsequently, the end goal of the BUPERS-34 model is to be able to better predict zone reenlistment rates. Adjusted R-square provides insight on how significant the factors, or variables, are fitted in the model. Unlike R-square, adjusted R-square adjusts for the number of model terms and increases only if the new term improves the model more than would be expected by chance.

2. Choice of Factors, Levels, and Range

Current forecasting procedures for reenlistment rate are broken into three zones: A, B, and C. As previously discussed, these three zones have separate retention models, broken down by zone (e.g., years in the navy), and are categorized by Zone A (0-6 years), Zone B (6-10 years), and Zone C (10-14 years). Each of the three zones has eight runs (experiments) for a total of 24 for this experimental design.

To review, the BUPERS-34 Reenlistment Rate Prediction model variables used to predict reenlistment rate by zone are applied in this DOX and are listed in Table 13.

Zone A variables	**Zone B variables**	**Zone C variables**
Unemployment Rate	Unemployment Rate	Unemployment Rate
Zone A Attrition Rate	Zone B Attrition Rate	Zone C Attrition Rate
Zone A Change in Fiscal Year End Strength	Fiscal Year Navy Enlisted End Strength	Fiscal Year Navy Enlisted End Strength For Years 10-13

Table 13. Design of Experiment Zone Variables

The purpose of this experimental design is to analyze the effect of the factors; amount of data, model type, and data frequency on standard deviation and adjusted R-square values. Table 14 lists these factors with their associated levels and data type.

Factor	**Levels**	**Modeling Type**
Amount of Data	5 year (-1) 10 year (+1)	Nominal
Model Type	Main Effects (-1) Two Factor Interaction (2FI) (+1)	Nominal
Data Frequency	Annual (-1) Monthly (+1)	Nominal

Table 14. Design of Experiment Factors and Levels

Design of experiments is used to determine the impact of the three factors (β_1 = amount of data, β_2 = model type, and β_3 = data frequency) on the two response variables which are the standard deviation and adjusted R-squared. The equations tested are:

$$Y_1 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_2 x_3 + \beta_{12} x_1 x_3 + \beta_{12} x_1 x_3 + \beta_{23} x_2 x_3 + \epsilon$$

$$Y_2 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_2 x_3 + \beta_{12} x_1 x_3 + \beta_{12} x_1 x_3 + \beta_{23} x_2 x_3 + \epsilon$$

where Y_1 is standard deviation and Y_2 is adjusted R-squared. For example if β_1 is significant for zone A, and the response variable is Y_1, then it will be concluded that amount of data has an impact on the standard deviation of the regression fit for zone A data.

3. Experimental Design

A 3-factor design with eight runs (2^3) for each zone is constructed using JMP 8, a statistical software package, and is depicted in Table 15. The design displays the coded units (-1, +1), which corresponds to the low (-1) and high (+1) levels for each variable. Refer to Montgomery (2008) for a detailed description of factorial design.

Run	Model Type (Main Effects/2FI)	Amount of Data (5/10 yr)	Rate of Data (Annual/Monthly)
1	1	1	1
2	1	-1	-1
3	-1	-1	1
4	-1	1	-1
5	-1	1	-1
6	1	1	1
7	1	-1	-1
8	-1	-1	1

Table 15. 3-Factor Experimental Design Randomization for Each Zone

4. Analyzing the Experiment

Eight runs are generated per zone for zones A, B, C. Table 16 records the standard deviation and adjusted R-square for each experiment.

The experiment excludes three runs because these runs contain less than the required degrees freedom in the two-factor interaction, resulting in insufficient data available to effectively analyze. Subsequently, Table 16 from JMP 8 depicts the results of 21 runs after removing the insufficient data.

	Model Type _Main Effects(-)/2FI(+)	Amount of Data _5(-)/10 (+) year	Rate of Data _Annual(-)/Monthly(	Standard Deviation (Y)	Adj. R^2	Zones A, B, or C
1	-	-	-	0.0191	0.8923	A
2	-	+	-	0.0167	0.8831	A
3	+	+	-	0.0181	0.8629	A
4	-	-	+	0.128	-0.0299	A
5	+	-	+	0.13	-0.0617	A
6	-	+	+	0.1148	0.04	A
7	+	+	+	0.1149	0.03	A
8	-	-	-	0.0103	0.9776	B
9	-	+	-	0.0365	0.6623	B
10	+	+	-	0.0438	0.5135	B
11	-	-	+	0.069	0.2708	B
12	+	-	+	0.0692	0.2654	B
13	-	+	+	0.0688	0.431	B
14	+	+	+	0.0667	0.4649	B
15	-	-	-	0.0007	0.9987	C
16	-	+	-	0.0119	0.7699	C
17	+	+	-	0.0129	0.7306	C
18	-	-	+	0.0355	0.1561	C
19	+	-	+	0.0352	0.1693	C
20	-	+	+	0.0359	0.2799	C
21	+	+	+	0.0361	0.2701	C

Table 16. Adjusted Design for the BUPERS-34 Reenlistment Rate Prediction Model

A row in the design matrix (first three and last column of Table 16) corresponds to a single experiment. As an example of how to conduct each experiment and collect response data, consider the first row in Table 16. Row 1 represents the results of an experiment using zone A data (see last column in Table 16). This experiment uses the low level of model type, low level of amount of data and low level for the rate of data. To run this experiment then, a reenlistment rate model is created for zone A using all three main effects (unemployment rate, attrition rate, and end strength) for the past five years, using yearly data. Once this model is created, the standard deviation and adjusted R-squared are recorded. In this experiment the standard deviation is 0.0191 and the adjusted R squared is 0.8923.

The results of the individual experiments are listed in Table 16. Approximately half of the adjusted R-square experiments, had values that are quite low, indicating that those particular experiments did not result in an adequate fit. In general, those experiments were generated with monthly data. Further investigation within the individual experiments indicates a possible seasonality trend within the data.

a. Statistical Analysis of the Response Variable- Standard Deviation

The initial linear regression (main effects) results for the DOX model with the dependent variable (Y_1) as standard deviation is depicted in Figure 6. As observed, the only significant effects are the variables “**Rate of Data**” and “**Zone.**”

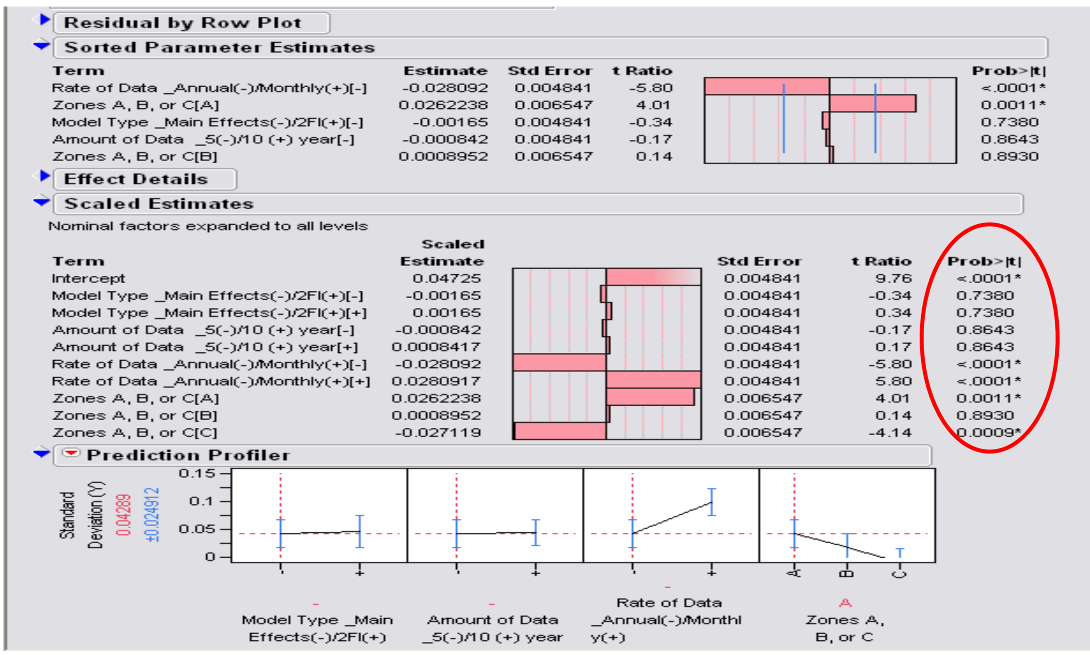
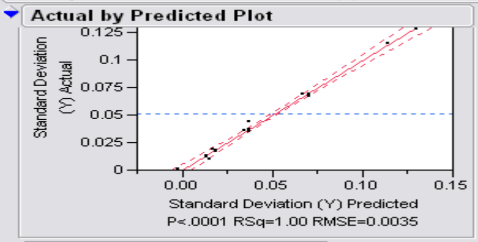


Figure 6. DOX Main Effects Results

Subsequently, a two-degree factorial and polynomial with stepwise linear regression is used, resulting with most variables having significance and an adjusted R-square of 99 percent (Figure 7).

Response Standard Deviation (Y)

Actual by Predicted Plot



Summary of Fit

RSquare	0.995972
RSquare Adj	0.992676
Root Mean Square Error	0.003486
Mean of Response	0.051148
Observations (or Sum Wgts)	21

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	9	0.03305358	0.003673	302.1768
Error	11	0.00013369	0.000012	Prob > F
C. Total	20	0.03318727		<.0001*

Lack Of Fit

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.0460083	0.000796	57.83	<.0001*
Amount of Data _5(-)/10 (+) year[-]	-0.002083	0.000796	-2.62	0.0239*
Rate of Data _Annual(-)/Monthly(+)[-]	-0.029333	0.000796	-36.87	<.0001*
Zones A, B, or C[A]	0.02379	0.001102	21.58	<.0001*
Zones A, B, or C[B]	0.001465	0.001102	1.33	0.2108
Amount of Data _5(-)/10 (+) year[-]*Rate of Data _Annual(-)/Monthly(+)[-]	-0.004558	0.000796	-5.73	0.0001*
Amount of Data _5(-)/10 (+) year[-]*Zones A, B, or C[A]	0.0057567	0.001102	5.22	0.0003*
Amount of Data _5(-)/10 (+) year[-]*Zones A, B, or C[B]	-0.004393	0.001102	-3.99	0.0021*
Rate of Data _Annual(-)/Monthly(+)[-]*Zones A, B, or C[A]	-0.022793	0.001102	-20.68	<.0001*
Rate of Data _Annual(-)/Monthly(+)[-]*Zones A, B, or C[B]	0.0083817	0.001102	7.60	<.0001*

Figure 7. DOX Two-Factor Interaction Results

It is observed from the interaction profile (Figure 8), resulting from the two-degree factorial and polynomial stepwise procedure, that **Rate of Data** and **Zone A, B, and C** has the greatest amount of variation (i.e., standard deviation effects), as well as, significance to standard deviation. As observed in the highlighted circle in Figure 8, as **Rate of Data** goes from annual to monthly, standard deviation varies significantly. This indicates that the level used for **Rate of Data** is significant to standard deviation. Because we have previously observed possible seasonal effects for monthly data (Table 16), and the level of **Rate of Data** is significant to Y_1, then the use of annual data in the model may be the best process to minimize standard deviation.

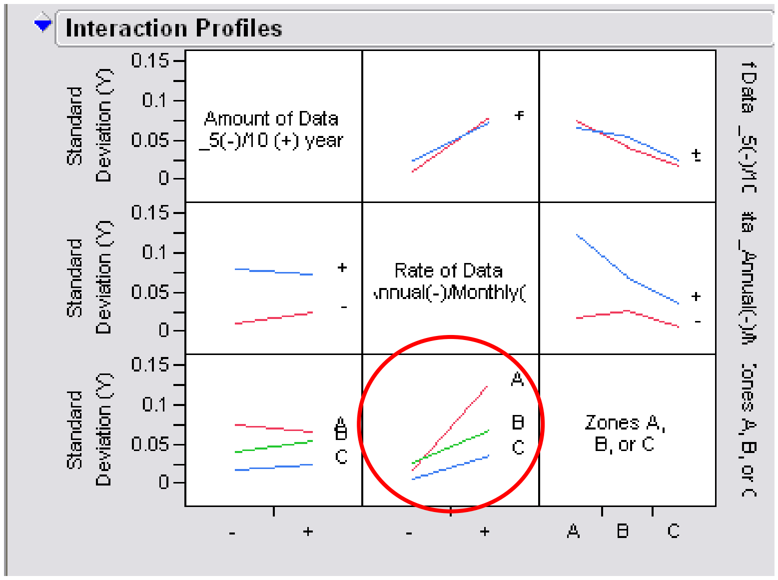
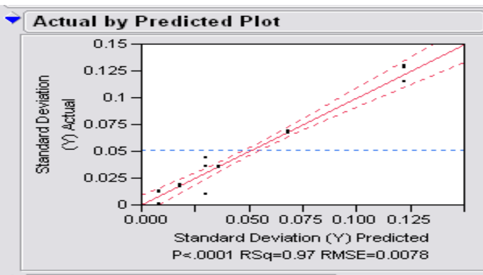


Figure 8. DOX Two-Factor Interaction Profiles

Subsequently, those two effects, **Rate of Data** and **Zone A, B, and C**, were isolated in a new model resulting in a high adjusted R-square of 97 percent, a small standard error of .007, and significant values for all but one variable as observed in Figure 9.



Summary of Fit

RSquare	0.972216
RSquare Adj	0.962954
Root Mean Square Error	0.00784
Mean of Response	0.051148
Observations (or Sum Wgts)	21

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	0.03226518	0.006453	104.9742
Error	15	0.00092209	0.000061	**Prob > F**
C. Total	20	0.03318727		<.0001*

Parameter Estimates

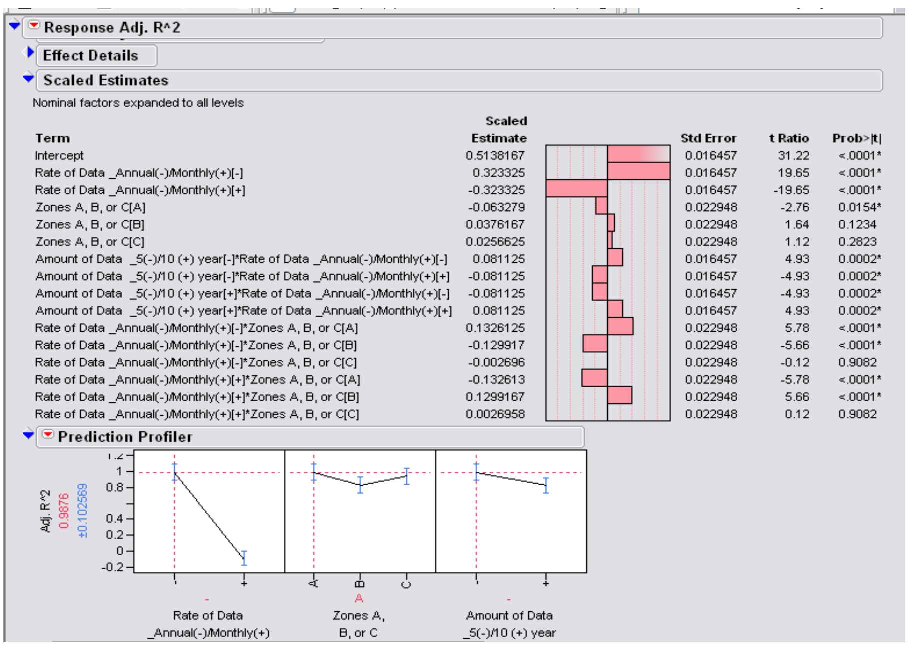
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.0471153	0.001729	27.26	<.0001*
Rate of Data _Annual(-)/Monthly(+)[-]	-0.028226	0.001729	-16.33	<.0001*
Zones A, B, or C[A]	0.0228306	0.002445	9.34	<.0001*
Zones A, B, or C[B]	0.0021972	0.002445	0.90	0.3830
Rate of Data _Annual(-)/Monthly(+)[-]*Zones A, B, or C[A]	-0.023753	0.002445	-9.72	<.0001*
Rate of Data _Annual(-)/Monthly(+)[-]*Zones A, B, or C[B]	0.0091139	0.002445	3.73	0.0020*

Figure 9. DOX Rate of Data and Zone A, B, and C Interaction Results

b. Statistical Analysis of Response Variable- Adjusted R-square

Referring to Table 16, as with standard deviation, there is a possible seasonality within the monthly data (when compared to the annual data), that results with low adjusted R-square values.

The initial linear regression (main effects) results for the DOX Adjusted R-square model are similar (in insignificance) to the results of those when the response variable Y_1. Subsequently, a stepwise fit with a two-degree factorial (2FI) and polynomial is used when the response variable is Y_2 with the results depicted in Figure 10.

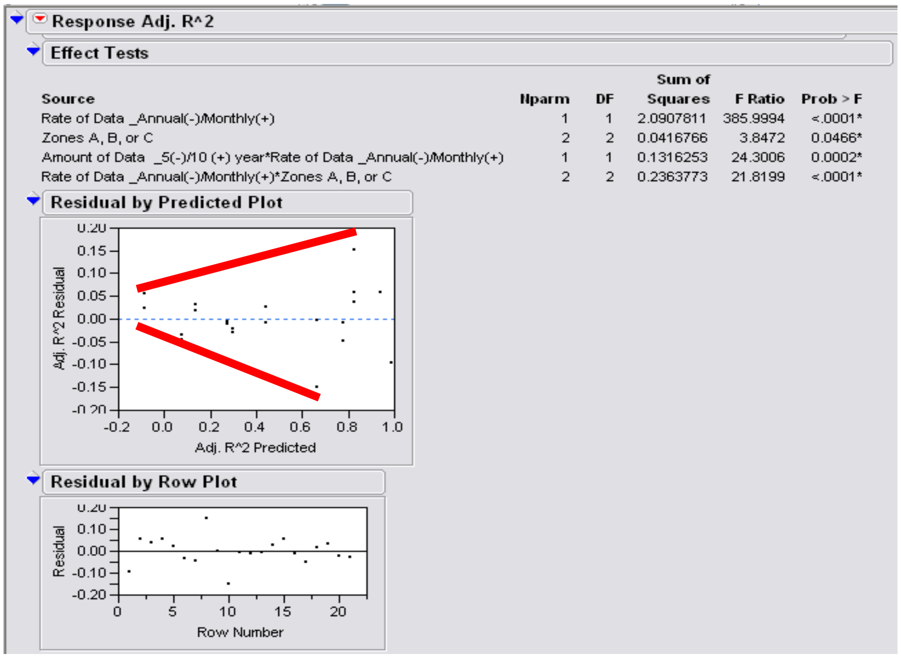


Term	Scaled Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.5138167	0.016457	31.22	<.0001*
Rate of Data _Annual(-)/Monthly(+)[-]	0.323325	0.016457	19.65	<.0001*
Rate of Data _Annual(-)/Monthly(+)[+]	-0.323325	0.016457	-19.65	<.0001*
Zones A, B, or C[A]	-0.063279	0.022948	-2.76	0.0154*
Zones A, B, or C[B]	0.0376167	0.022948	1.64	0.1234
Zones A, B, or C[C]	0.0256625	0.022948	1.12	0.2823
Amount of Data _5(-)/10 (+) year[-]*Rate of Data _Annual(-)/Monthly(+)[-]	0.081125	0.016457	4.93	0.0002*
Amount of Data _5(-)/10 (+) year[-]*Rate of Data _Annual(-)/Monthly(+)[+]	-0.081125	0.016457	-4.93	0.0002*
Amount of Data _5(-)/10 (+) year[+]*Rate of Data _Annual(-)/Monthly(+)[-]	-0.081125	0.016457	-4.93	0.0002*
Amount of Data _5(-)/10 (+) year[+]*Rate of Data _Annual(-)/Monthly(+)[+]	0.081125	0.016457	4.93	0.0002*
Rate of Data _Annual(-)/Monthly(+)[-]*Zones A, B, or C[A]	0.1326125	0.022948	5.78	<.0001*
Rate of Data _Annual(-)/Monthly(+)[-]*Zones A, B, or C[B]	-0.129917	0.022948	-5.66	<.0001*
Rate of Data _Annual(-)/Monthly(+)[-]*Zones A, B, or C[C]	-0.002696	0.022948	-0.12	0.9082
Rate of Data _Annual(-)/Monthly(+)[+]*Zones A, B, or C[A]	-0.132613	0.022948	-5.78	<.0001*
Rate of Data _Annual(-)/Monthly(+)[+]*Zones A, B, or C[B]	0.1299167	0.022948	5.66	<.0001*
Rate of Data _Annual(-)/Monthly(+)[+]*Zones A, B, or C[C]	0.0026958	0.022948	0.12	0.9082

Figure 10. DOX Adjusted R-square Rate of Data and Zone Interaction Results

The results from the two-degree factorial and polynomial linear regression show that **Rate of Data** and **Amount of Data** are most significant to adjusted R-square (Figure 10). In examining the residuals, which are estimates of experimental error obtained by subtracting the observed responses from the predicted responses, it is observed in the Residual by Predicted Plot (in Figure

11) that the residuals form a funnel shape. This may indicate that transforming the response variable is required, or may indicate that the data points (i.e., random variables) are not NID, and, subsequently, may not have the same probability distribution and not be statistically independent.



Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Rate of Data _Annual(-)/Monthly(+)	1	1	2.0907811	385.9994	<.0001*
Zones A, B, or C	2	2	0.0416766	3.8472	0.0466*
Amount of Data _5(-)/10 (+) year*Rate of Data _Annual(-)/Monthly(+)	1	1	0.1316253	24.3006	0.0002*
Rate of Data _Annual(-)/Monthly(+)*Zones A, B, or C	2	2	0.2363773	21.8199	<.0001*

Figure 11. DOX Adjusted R-square Residual Plot For Rate Data and Zone Interaction

5. Experimental Design Insights

This design of experiments is used to analyze the fit of the current BUPERS-34 Reenlistment Rate Prediction method and improve the performance of the model and/or discover insights that can be used to develop alternative models to improve reenlistment rate predictions. **Rate of Data** with the interaction of **Zone A, B, C** are significant with the dependent variable, Y_1. The factors, **Rate of Data** and **Amount of Data** appeared to be significant with the dependent variable, Y_2.

The values in Table 16 consistently indicate that the level, monthly data, within the factor, Rate of Data, results in poor adjusted R-square values and large standard deviation values relative to all other model factors and levels.

The following summarizes the DOX insights:

1. Fiscal Year (i.e., annual) data produces lower standard deviations and higher adjusted R-square values in regression analysis.
2. Two-factor interaction does not improve performance.
3. There is a significant interaction between **Zone** and **Rate of Data** (i.e., monthly, annual). This indicates that the use of annual data over monthly data may lead to a more robust model because it minimizes variation in measuring retention effects for zones A, B, and C.
4. 10-year fiscal data produces more significant results than 5-year fiscal data due to the 5-year fiscal year data having insufficient amount of degrees of freedom. There is not a sufficient amount of historical data to conduct an experiment for a 15-year or greater period.
5. Zone A 10-year annual data produces the best adjusted R-square and standard deviation results (with significant p-values) over the 21 experiments followed by zone C and then zone B.
6. 5-year annual data also produces excellent adjusted R-square values with low standard deviation values. However, upon review of the prediction variable p-values from these experiments, some are found to be insignificant which would lead to a poor model fit.

THIS PAGE INTENTIONALLY LEFT BLANK

V. DEVELOPING A NEW MODEL FOR PREDICTING REENLISTMENT RATE

The analysis gained from the study of related literature in Chapter II, reviewing and examining of the BUPERS-34 Reenlistment Rate Prediction model methodology in Chapter III, and insights from Chapter IV's experimental design provide great direction in developing a new prediction model for BUPERS- 34 to predict reenlistment rates by zone in the aggregate.

This chapter presents several alternative mathematical models that can be used for predicting future reenlistment rates. The goal of the alternative models is to improve both accuracy and precision in the predictions made. Improving accuracy means that the Navy will have a better idea what the true reenlistment rates will be and improving precision equates to reduced prediction variance.

Several alternative options for predicting reenlistment rate are investigated. Time series analysis is suggested to deal with the violation of assumptions in the linear regression model and the addition of a variable—"**SRB**" —is suggested as an improvement to the model.

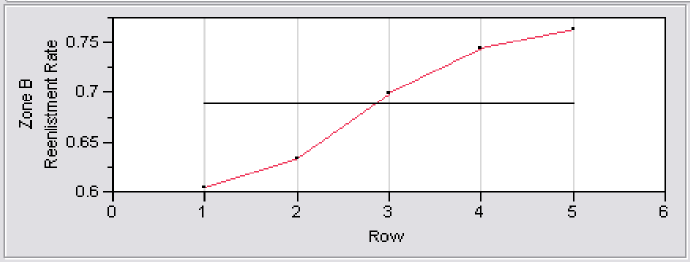
A. TIME SERIES EXPERIMENT

A time series experiment is conducted to analyze and forecast reenlistment rate annual data. Because a time series is a set of observations $\{y_1, y_2, \ldots, y_n\}$ taken over a series of equally-spaced time periods, as is the case with the reenlistment rate annual data, this experiment is of value to investigate and determine the strength of time series forecasting on reenlistment rate annual data.

In this time series experiment, an Integrated Moving Average (IMA) model is selected which predicts future values of a time series by a linear combination of its past values and a series of errors (also known as random shocks or innovations). The IMA model used is equivalent to the exponentially weighted moving average (EWMA) technique.

Figure 12 is a time series IMA model. Displayed are autocorrelations and partial autocorrelations of the BUPERS-34 5 year zone B data modeled in time series. These indicate how and to what degree each point in the series is correlated with earlier values in the series. The IMA is selected as the best specified Autoregressive Integrated Moving Average (ARIMA) model to perform a maximum likelihood fit of the data to time series for the 5 year zone B reenlistment rates, resulting with an adjusted R-square of 87 percent. For consistency of the experiment, the IMA model is used to determine the adjusted R-square and standard deviations of each run.

Time Series Zone B Reenlistment Rate



Mean	0.6892
Std	0.0617297
N	5
Zero Mean ADF	3.4634603
Single Mean ADF	-0.617552
Trend ADF	-1.381156

Time Series Basic Diagnostics

Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	0.4672	1.9101	0.1669	1	0.4672
2	-0.1694	2.2450	0.3255	2	-0.4960
3	-0.4672	6.0654	0.1085	3	-0.1986
4	-0.3306	9.8899	0.0423*	4	-0.0313

Model Comparison

Model	DF	Variance	AIC	SBC	RSquare	-2LogLH	AIC Rank	SBC Rank	MAPE	MAE
IMA(1, 1)	2	0.0002814	-18.51458	-19.74199	0.918	-22.51458	0	0	2.052323	0.014458

Model: IMA(1, 1)

Model Summary

DF	2	Stable	Yes
Sum of Squared Errors	0.00056283	Invertible	Yes
Variance Estimate	0.00028142		
Standard Deviation	0.01677548		
Akaike's 'A' Information Criterion	-18.514576		
Schwarz's Bayesian Criterion	-19.741987		
RSquare	0.91766771		
RSquare Adj	0.87650157		
MAPE	2.05232292		
MAE	0.01445833		
-2LogLikelihood	-22.514576		

Figure 12. Zone B, 5-Year, Annual, IMA Time Series Model

The adjusted R-square and standard deviation for each time series IMA (1,1) run is recorded in Table 17 and compared to their respective values derived earlier (Table 16) using regression analysis. Two runs for each zone (e.g., Amount of Data, five and ten year) are conducted.

Run	Amount of Data 5 yr (-) 10yr (+)	Rate of Data Annual(-) Monthly(+)	Table 17 Standard Deviation (Y)	Table 17 Adjusted R-square	Zones A, B,C	Adjusted R-square Annual Time Series IMA(1,1)	Standard Deviation Annual Time Series IMA (1,1)
1	-	-	0.0191	0.8923	A	-0.1221	0.0369
2	+	-	0.0167	0.8831	A	-0.4571	0.0573
3	-	-	0.0103	0.9776	B	0.8765	0.0167
4	+	-	0.0365	0.6623	B	0.3318	0.0524
5	-	-	0.0007	0.9987	C	0.1514	0.0141
6	+	-	0.0119	0.7699	C	0.2557	0.0218

Table 17. Time Series Design for the BUPERS-34 Enlisted Retention

A least squares regression is then performed from the adjusted R-square values (in Table 17). Most of the runs show minimal significance (e.g., ability to predict future outcomes) to the model as indicated by the p-values in Figure 13, resulting in an adjusted R-square of 72 percent, and indicate that the individual model runs (in Table 17) may not be good predictors for future outcomes.

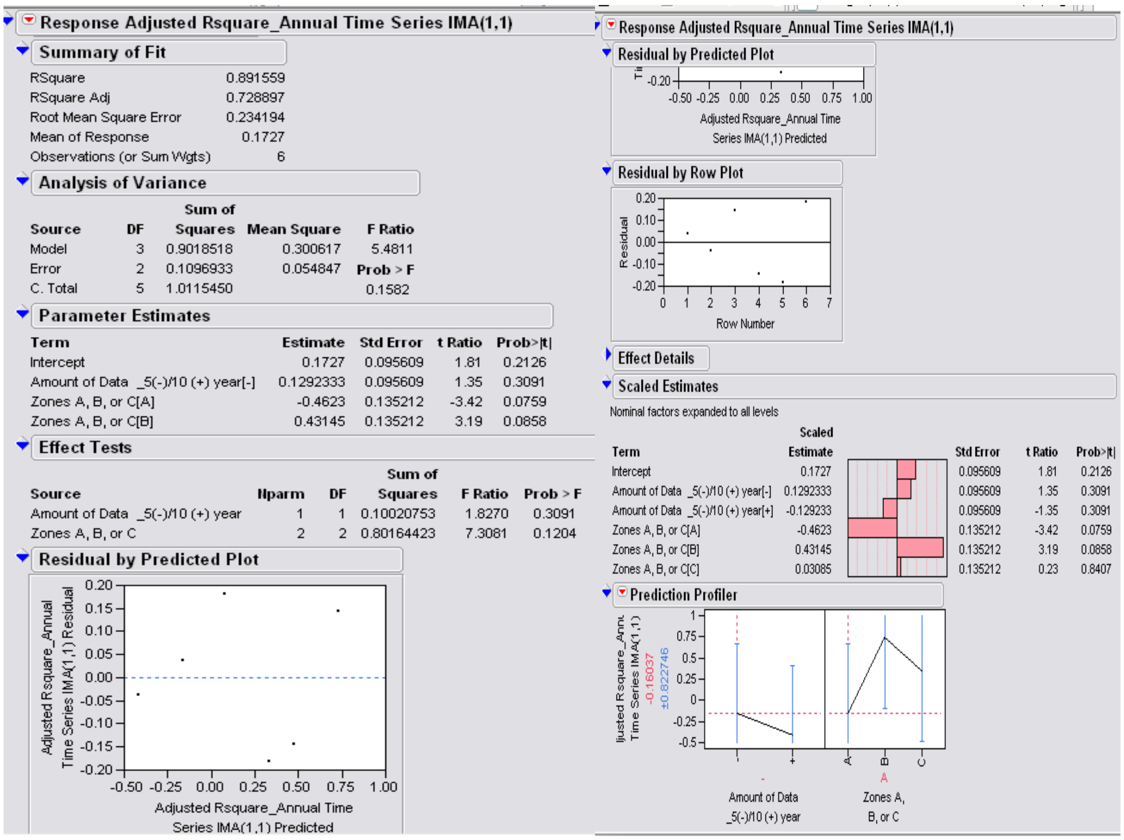


Figure 13. Time Series IMA Regression Analysis on Adjusted R-square

A least squares regression is performed for the standard deviation values resulting in an adjusted R-square of 71 percent. There are no significant variables in the results captured in Figure 14.

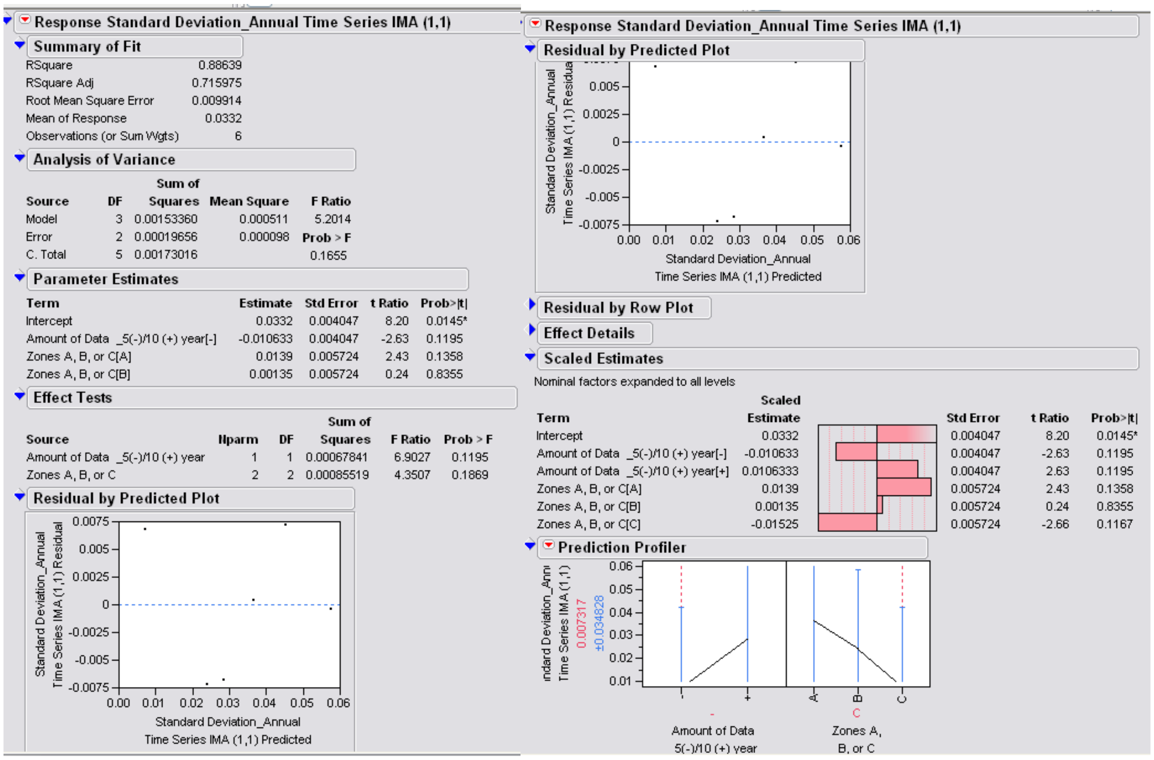


Figure 14. Time Series IMA Regression Analysis on Standard Deviation

1. Time Series Experimental Design Insights

The time series IMA experiment does not yield any significant information other than this particular ARIMA model does not produce a significant improvement for predicting reenlistment rates.

As presented in Chapter IV, the prediction variable, “**Unemployment Rate,**” in the current linear model, is not significant when combined with the other model variables. This section explores removing the insignificant model variables and adds a variable representing “**SRB,**” which is suggested as a significant variable in the literature. Additionally, as observed in Chapter IV’s experimental design, the data is varied over different periods to investigate significance.

Before dropping insignificant variables from regression analysis in the following models, this thesis defines a variable to be an insignificant variable if the p-value is less than the significance level of 0.05. In probability theory, this is the acceptable level of a Type I error; it is the risk of falsely rejecting the null hypothesis. Subsequently, lower p-values mean lower probabilities of committing Type I errors. Additionally, variable selection plays a critical role in determining the relevance of a prediction variable on a response variable (e.g., Reenlistment Rate). For example, BUPERS-34 uses total navy end strength to predict zone B reenlistment rates. However, zone B end strength, a much smaller and specific subset of total navy end strength may be a more relevant and appropriate variable to measure the zone B reenlistment rate.

Model variables are investigated in detail and additional variables are researched to include various unemployment rates acquired from the BLS, Consumer Confidence data, various end strength calculations, aggregate pay increases, reenlistment programs, and SRB data. From the literature review, SRB data is found to be significant in enlisted retention. SRB is a significant lever in Navy enlisted retention because it is easily modified and can be continuously adjusted to meet retention targets. Each zone is analyzed to see if SRB is significant to the model. Additionally, several models are developed over various time periods to analyze and provide the statistical variation necessary to produce significant estimates to predict the reenlistment rate.

Stepwise multivariate regression analysis is conducted on zones A, B, and C. Results are found in Table 18.

Zones	**Model Fit**	**Dependent Variables**			
	Adjusted R-square	**End Strength**	**Unemployment Rate**	**Attrition Rate**	**Zone SRB**
Zone A	**0.916**	.0005*		.0026*	.0439*
Zone B	**0.869**	.0046*	.0256*	.0018*	
Zone C	**0.749**			.0004*	

* P Value Significance at 0.05

Table 18. New Reenlistment Rate Prediction Model Fits

The resulting Multiple Linear Regression response and predictor variables for zones A, B, and C are summarized in Table 19.

Variable	Variable Description
Zone A	
y	**Reenlistment Rate**. Reenlistment rate data from the previous 10 FYs is obtained from NRMS.
x_1	**End Strength**. Change in zone A end-strength from the previous 10 FYs is obtained from NRMS.
x_2	**Attrition Rate**. Attrition rate data from the previous 10 FYs is obtained from NRMS.
x_3	**Zone SRB**. Fiscal year zone A SRB totals from the previous 10FYs is obtained from N13.
Zone B	
y	**Reenlistment Rate**. Reenlistment rate data from the previous 10 FYs is obtained from NRMS.
x_1	**End Strength**. Zone B end-strength at the end of the FY for previous 10 FY's is obtained from NRMS.
x_2	**Unemployment Rate**. Unemployment rate data from the previous 10 calendar years (CY) is obtained from the Bureau of Labor Statistics.
x_3	**Attrition Rate**. Attrition rate data from the previous 10 FYs is obtained from NRMS.
Zone C	
y	**Reenlistment Rate**. Reenlistment rate data from the previous 11 FYs is obtained from NRMS.
x_1	**Attrition Rate**. Attrition rate data from the previous 11 FYs is obtained from NRMS.

Table 19. Zones A, B, and C Response and Predictor Variables

2. Zone A Alternative Reenlistment Rate Prediction Model

Zone A SRB is found significant to predicting zone A reenlistment rates. By including **Zone A SRB** for the last 10 FYs and removing **Unemployment Rate**, which had a p-value of .268 in the BUPERS-34 prediction model, as a predictive variable for zone A, adjusted R-square improved in the Reenlistment Rate Prediction model from .883 to .916 (Figure 15). Residuals appear to be NID (0, σ^2), and the removal of insignificant variables resolves the over dispersion problems that existed in the BUPERS-34 Zone A prediction model.

From the results in Figure 15, the fitted regression equation can be written as:

$$Y_{\text{zone A RE Rate}} = 0.634 + 0.0000057x_1 - 1.76555x_2 + .000000006x_3$$

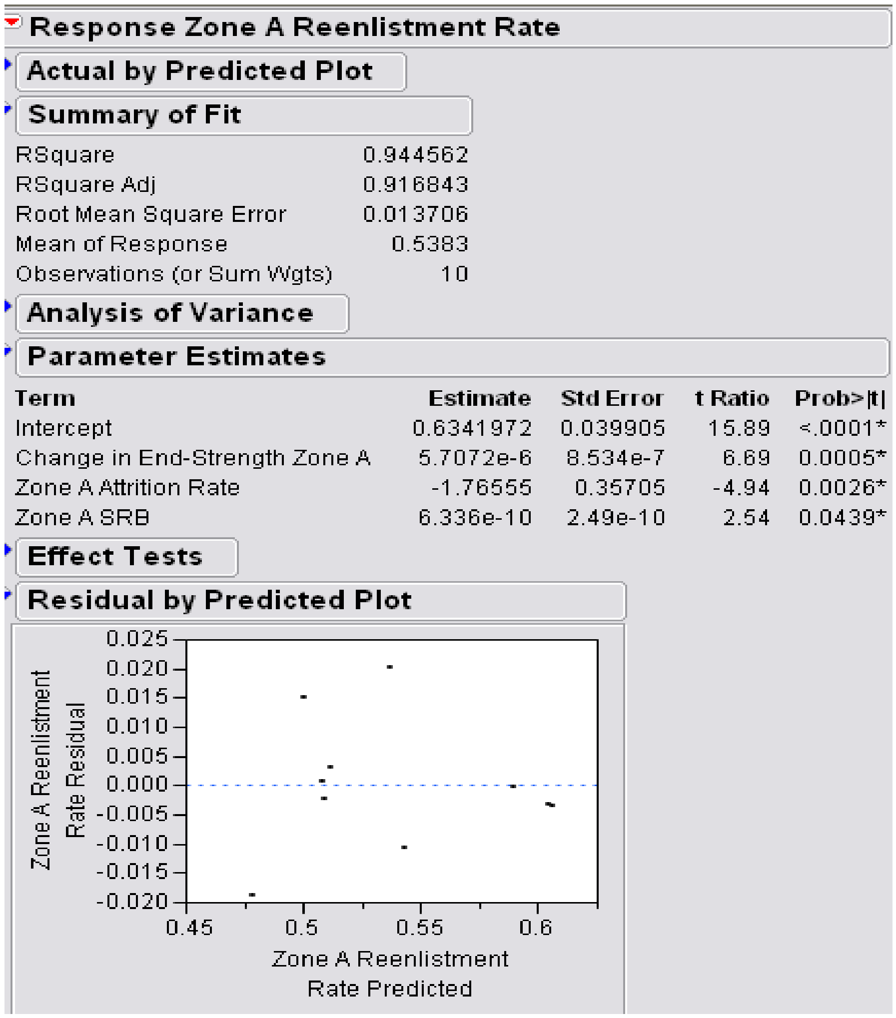


Figure 15. Zone A Alternative Model Regression Analysis

3. Zone B Alternative Reenlistment Rate Prediction Model Results

The unemployment rate is found significant to predicting zone B reenlistment rates using the last 10 FYs. Goldberg says by reducing the period of the prediction model from the last 15 FYs (as in the BUPERS-34 model) to the last 10 FYs, the statistical variation (as seen with unemployment rates) necessary to produce significant estimates to predict the reenlistment rate significantly improves (Goldberg, 1986). Additionally, a new variable, "**Zone B End Strength,**" replaces the BUPERS-34 end strength variable which measures "**Total Navy End Strength**" used in the prediction model. Consequently, adjusted R-square significantly improves in the Reenlistment Rate Prediction model from .461 to .869 (Figure 16), greatly increasing the model's prediction capability. Residuals appear to be NID $(0, \sigma^2)$, and the removal of insignificant variables and adjusting the period to the last 10 FYs, resolves the over dispersion problems that exist in the BUPERS-34 zone B prediction model, and significantly increases the model's predictive capability.

From the results in Figure 16, the fitted regression equation can be written as:

$$Y_{\text{zone B RE Rate}} = 0.9951 - 0.0000066x_1 + 1.70959x_2 - 6.77031x_3$$

Response Zone B Reenlistment Rate

Actual by Predicted Plot

Summary of Fit

RSquare	0.912674
RSquare Adj	0.869011
Root Mean Square Error	0.02326
Mean of Response	0.656
Observations (or Sum Wgts)	10

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	0.03392774	0.011309	20.9027
Error	6	0.00324626	0.000541	Prob > F
C. Total	9	0.03717400		0.0014*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.9951016	0.078106	12.74	<.0001*
BLS Monthly Unemployment Rate	1.7095941	0.579494	2.95	0.0256*
Zone B Attrition Rate	-6.77031	1.278518	-5.30	0.0018*
Zone B End Strength	-6.602e-6	1.504e-6	-4.39	0.0046*

Effect Tests

Residual by Predicted Plot

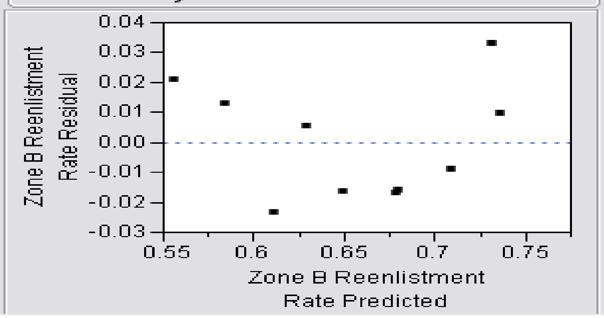


Figure 16. Zone B Regression Analysis

4. Zone C Alternative Reenlistment Rate Prediction Model Results

Zone C Attrition Rate explains 75 percent of the variability in predicting zone C reenlistment rates over the last 11 FYs. Removing **Unemployment Rate** as a predictive variable for zone C results in no significant change to the model's fit with adjusted R-square remaining nearly the same as the BUPERS-34 prediction model (Figure 17). The removal of **Unemployment Rate** as a prediction variable resolves the over dispersion problems that existed in the BUPERS-34 zone C prediction model because it does not explain any of the variability in the model and is not statistically significant.

From the results in Figure 17, the fitted regression equation, with only attrition rate as an input, can be written as:

$$Y_{\text{zone C RE Rate}} = 0.887 - 4.4554x_1$$

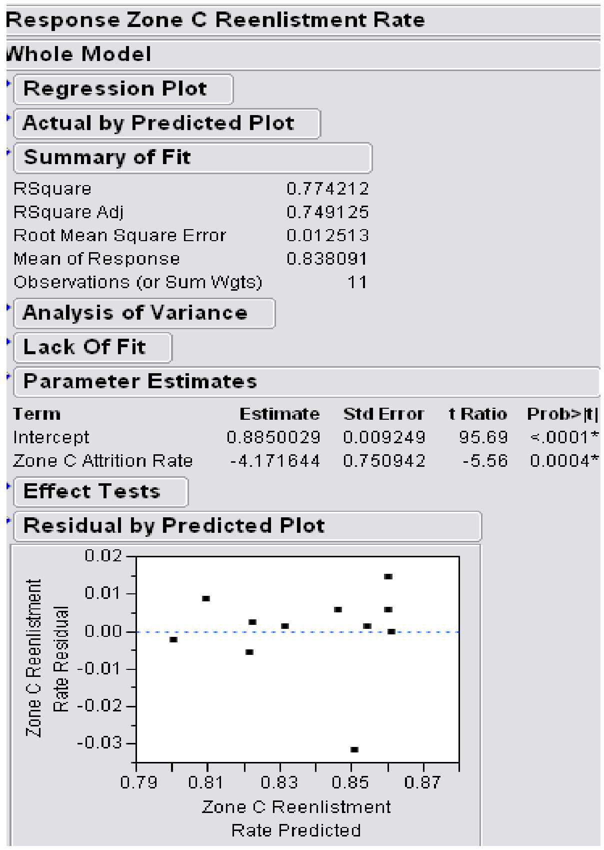


Figure 17. Zone C Regression Analysis

VI. CONCLUSIONS

A. SUMMARY

BUPERS-34 predicts zone A, B, and C reenlistment rates using multivariate linear regression within Excel. However, the BUPERS-34 Reenlistment Rate Prediction model used to predict reenlistment rates for Navy FY09 and FY10 retention goals has three main problems; it violates the assumption that the residual errors are NID $(0, \sigma^2)$, it uses insignificant variables and/or inferior variable selection, and some prediction variables require predictions in order to make forecasts for future values. Additionally, model variables are never investigated for 2FI by BUPERS-34.

This thesis uses several statistical techniques available within the statistical software JMP 8 and recommends an alternative model to each of the three zones, A, B, and C, that is more robust than the current BUPERS-34 prediction models. The alternative models eliminate insignificant variables and/or inferior variable selection, improve model robustness and model fit for all zones, and investigate and incorporate additional compensation and non-compensation variables that effect zone reenlistment rate predictions. All of which lead to improved prediction capabilities. Table 20 provides a comparison between the BUPERS-34 adjusted R-squared values and the proposed model adjusted R-squared values. While the adjusted R-squared for zone C is slightly decreased, the model is considered improved because of the removal of insignificant variables, which add noise to the predictions.

Zone	BUPERS-34 Adjusted R-square	Alternative Model Adjusted R-square
Zone A	.883	.916
Zone B	.461	.869
Zone C	.756	.749

Table 20. Zones A, B, and C Model Fit Comparisons

The recommended models are still regression models. These models only use ten years of historical data and do not appear to violate the residual assumptions. Further work should investigate using a time series in conjunction with regression analysis. In addition, while these alternative models may be the best for this year, it is recommended that each zone model be updated, reevaluated, and checked for significance and fit on an annual basis.

B. FUTURE RECOMMENDATIONS AND RESEARCH

1. Total Force Database

Retention measures (e.g., reenlistment rates and attrition rates) and other retention variables are stored and calculated in the NRMS, Navy's authoritative source of retention. NRMS is the primary data source used to provide timely and accurate reporting and analysis of reenlistment, retention, and attrition data to the Fleet. However, NRMS has several drawbacks. For example, SRB, a dimension within NRMS and a significant variable within the alternative prediction model for zone A, is not reliably populated due to limited resources and/or funding. Some calculations are inconsistent. Policy guidance mandates that retention calculations are to be standardized; however, end strength calculations differ between N100 and BUPERS-34 depending on if calculation is used towards retention or towards end strength forecasts (i.e., N100 includes short term extensions in RE denominator) (Chilson, 2009).

The requirement for BUPERS-34 and other analysts to develop prediction models and/or forecasts requires empirical data that is not always available. The need for one standardized tool that is designed to provide researchers with ready access to the personnel, manpower and related data required for empirical analysis of retention, enlistment, and other types of behavior that are of policy interest to the Navy is critical.

2. Aggregate Modeling

BUPERS-34 is required to predict zone reenlistment rates and numerical totals for the out-year FY retention goals. Some of the historical data available is constrained to shorter periods that lead to poor models due to insufficient degrees of freedom, or questionable results due to minimal data points.

Further research is recommended using Time Series analysis to model reenlistment rate behavior. As observed in Chapter IV, seasonality within the monthly data indicates that the residual errors are not NID (0, σ2). Further investigation using various seasonality models may result in improving the predictability for the zone reenlistment rates.

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Box, G. E. P., & Draper, N. R. (1987). *Empirical model-building and response surfaces*. New York: Wiley.

Budding, R., Levy, D. S., Hanley, J. M., & Waldman, D. M. (1992). *Promotion tempo and enlisted retention*. Rand Corporation.

Congressional Budget Office (CBO). (1981, July). *CBO Military Retention Model*, Congress of the United States Congressional Budget Office.

Congressional Budget Office (CBO). (2007, June). *Evaluating Military Compensation*. Pub. No. 2665. Congress of the United States Congressional Budget Office.

Chief of Naval Operations Organization Chart. (2007). [Chart of the Office of the Chief of Naval Operations on January 10, 2007]. *Navy Organization*. Retrieved on September 2010, from http://www.navy.mil/navydata/organization/org-cno

Chilson, C. (2009). *FY10 Enlisted Retention Goals* [PowerPoint slides], 2009.

Defense Technical Information Center (DTIC), *Title IV- Military Personnel Authorizations.* Retrieved on September 2010, from http://www.dtic.mil/congressional_budget/pdfs/FY2010_pdfs/HASC_111-166_MILPERS.pdf

Faram, M. (2010, May 31). SRB level reset for 3rd time in fiscal' 10. *Navy Times*, pp. 30.

Ferguson, M. E. (2008a). *Standard Enlisted Retention Measures*, Navy Administrative Message, No. 333/08.

Ferguson, M. E. (2008b). *Fiscal Year 2009 Retention Goals*, Navy Administrative Message, No. 348/08.

Ferguson, M. E. (2009). *Fiscal Year 2010 Retention Goals*, Navy Administrative Message, No. 321/09.

Goldberg, M. S. (2001). *A survey of enlisted retention: Models and findings*. Center for Naval Analyses.

Goldberg, M. S., & Hagar, M. F. (1981, June). *A comparison of the Prophet and ACOL Force Projection Model*. Professional Paper, p. 384.

Goldberg, M. S. (1986, October). *New estimates of the effect of unemployment on enlisted retention*. Center for Naval Analyses.

Hansen, M. L., & Wenger, J. W. (2003, October). *Is retention too high?* Center for Naval Analyses.

Montgomery, D. C. (2008). *Design and analysis of experiments*. New Jersey: Wiley.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2006). *Introduction to linear regression analysis*. New Jersey: Wiley.

Navy Organization. [Organizational chart, January 07, 2007] *Office of the Chief of Naval Operations*. Retrieved on September 2010, from http://www.navy.mil/navydata/organization/org-cno.asp

Questor, A. O., Hattaingadi, A. U., & Shuford, R. W. (2006, January). *Marine Corps retention in the post 9/11 era: The effects of deployment tempo on Marines with or without dependents.* Center for Naval Analyses.

Space and Naval Warfare (SPAWAR). (2004, August). *Navy Retention Monitoring System (NRMS) User Manual*.

Trumble, D., & Flanagan, D. M. (1990, December). *Feasibility study to update the Annualized Cost of Leaving (ACOL) procedures and the Naval Personnel Research and Development Center (NPRDC).* Oak Ridge National Laboratory.

United States Army Research Institute for the Behavioral Sciences. (2005, June). *A model of reenlistment behavior: Estimates of the Army's selective reenlistment bonus on retention by occupation* (Study Report 2005-02). Arlington, VA.

United States Department of Labor, Bureau of Labor Statistics. (2010). *Unemployment Rate (16 years and over)* [Data file]. Retrieved on September 2010, from http://www.bls.gov/data/unemployment.

United States House of Representatives. Office of the Law Revision Counsel. United States Code Title 10. Retrieved on September 2010, from http://uscode.house.gov/download/pls/10C2.txt

Warner, J. T. (1981, August). *Military compensation and retention: An analysis of alternative models and a simulation of a new retention model*. Center for Naval Analyses.

Welgan, M. (2010). *BUPERS-34 Metrics and Analytics Division Retention Responsibilities* (PowerPoint slides), 2010.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Rachel T Johnson
   Naval Postgraduate School
   Monterey, California

4. BUPERS-3
   Millington, Tennessee

5. BUPERS-34
   Millington, Tennessee